INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



MONUMENTOS MEGALITICOS

Septiembre 1980

250 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 48
Septiembre 1980

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

INVESTIGACION Y CIENCIA

DIRECTOR Francisco Gracia Guillén
REDACCION José María Valderas Gallardo (Redactor Jefe)
Carlos Oppenheimer
Encarna de Blas (Secretaria de Redacción)
PRODUCCION César Redondo Zayas
PROMOCION Y PUBLICIDAD Elena Sánchez-Fabres
PROMOCION EXTERIOR Pedro Clotas Cierco
EDITA Prensa Científica S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

Cristóbal Mezquita: *Lupus eritematoso sistémico;* Manuel Puigcerver: *Astronomía de líneas de rayos gamma;* Antonio Blanco: *Monumentos megalíticos;* José Luis Castillo: *Convección;* Joandomènec Ros: *Fisiología del koala;* Enrique Cerdá: *ADN superenrollado;* Juan José Martínez: *La hélice;* Luis Bou: *Juegos matemáticos;* J. Vilardell: *Taller y laboratorio.*

Libros:
Arnau, Ramón Margalef y Enrique Macpherson



LA PORTADA

En la fotografía de la portada se ven tres de las 12 piedras que tuvo originariamente el monumento megalítico de Stenness, en la isla de Orkney. Este conjunto megalítico mide algo más de los 30 metros de diámetro y está rodeado de un foso de algo más de 40 metros de radio, producido por la extracción de unas 1300 toneladas de arenisca. A diferencia de las piedras de Stonehenge, el más famoso de los monumentos megalíticos anulares, a las piedras de Stenness no se les dio una forma regular después de su extracción de la cantera. Las cuatro que permanecen erguidas tienen todas ellas un tamaño casi igual: 4,80 metros. El análisis del carbono 14 de muestras de madera y hueso asociadas con el monumento indica que fue erigido a fines del tercer milenio a. de C., o poco antes de la segunda fase de construcción de Stonehenge, en el Neolítico tardío (véase "Monumentos megalíticos", por Glyn Daniel, en este mismo número).

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2.750 pesetas
Extranjero:
Un año (12 números): 43 U.S.$
Ejemplar atrasado ordinario:
280 pesetas
Ejemplar atrasado extraordinario:
420 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49, Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239, Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137, Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.º, 2.ª, Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Evan Hadingham

Página	Fuente
9-13	M. Alfonseca, M.ª L. Tavera y M. Alonso
15	David Koffler, Hahnemann Medical College
16-20	Bunji Tagawa
22	David Koffler, Hahnemann Medical College
23-24	Bunji Tagawa
27	Marvin Leventhal, Bell Laboratories
28-34	Ilil Arbel
43	Cliché Lapie-Photothèque (*arriba*), West Air Photography (*abajo*)
44	Patricia J. Wynne
45	Evan Hadingham
46	Photograph© Françoise Legrand
47	Patricia J. Wynne
48	Roger Viollet (*arriba*), Irish Tourist Board (*abajo*)
49-51	Evan Hadingham
52	West Air Photography
54	Clemente Simón, Spainfo Ingenieros, S.A.
56-64	Alan D. Iselin
65	H. Linde, Instituto Central de Química Física, Berlín
66-67	Alan D. Iselin
71	Keith Gillett, Animals, Animals
72-75	Tom Prentiss
79	Gary Cohen, Universidad estatal de Nueva York en Stony Brook
80-88	George V. Kelvin
89	Jerome Vinograd, Instituto de Tecnología de California
91	Elsam, Mann and Cooper, Ltd. (*arriba*), Ralph Morse (*abajo*)
92-94	Dan Todd
95	F. M. N. Brown, Universidad de Notre Dame (*arriba*); Dan Todd (*abajo*)
96	Steven Finberg
97	Don Monroe
98	Dan Todd
99	Justin E. Kerwin, Instituto de Tecnología de Massachusetts
100	Lewis Research Center (NASA)
102-108	Gabor Kiss
111-117	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Imprime Grafesa. Nápoles, 249
Barcelona-13
Printed in Spain - Impreso en España

Los autores

MANUEL ALFONSECA y MARIA LUISA TAVERA ("La traducción de los lenguajes de computador") son miembros del Centro de Investigación UAM-IBM de Madrid. Alfonseca cursó estudios de ingeniería de telecomunicación en Madrid, y se doctoró en 1971 por la Universidad Politécnica de Madrid. Es también licenciado en informática, y, desde 1977, profesor encargado de lenguajes formales en la Facultad de Informática de la Universidad Politécnica de Madrid. En 1972 ingresó en el Centro de Investigación de IBM, donde ha realizado diversos trabajos relacionados con traducción de lenguajes, simulación y gráficos por computador. Tavera cursó estudios de ingeniería industrial en Madrid y se doctoró por la Universidad Politécnica de Madrid en 1980. También es licenciada en informática por la Universidad de Londres.

DAVID KOFFLER ("Lupus eritematoso sistémico") es profesor de patología y director del laboratorio de medicina del Hospital y Facultad de Medicina Hahnemann de Filadelfia. Es también profesor adjunto de inmunología de la Universidad Rockefeller. Inició estudios de derecho en la Universidad de Nueva York y ulteriormente estudió medicina, licenciándose por la Universidad estatal de Nueva York en 1958. Refiere el autor que la influencia más profunda en su carrera de investigador la recibió del inmunólogo Henry G. Kunkel, en cuyo laboratorio de la Universidad Rockefeller se inició en la investigación y con quien ha continuado colaborando en los últimos diez años.

MARTIN LEVENTHAL y CRAWFORD J. MACCALLUM ("Astronomía de líneas de rayos gamma") son físicos con un interés común por el estudio espectroscópico de procesos celestes que emiten radiación gamma a longitudes de onda discretas. Leventhal es miembro del personal técnico de los Laboratorios Bell. Estudió física en el City College de Nueva York y en la Universidad Brown, obteniendo su doctorado por esta última en 1964. Formó parte del cuerpo docente de la Universidad de Yale hasta 1968, en que ingresó en los Laborarios Bell. Los principales campos profesionales de interés de Leventhal, además de la astrofísica de rayos gamma, incluyen la investigación de cuerpos sólidos amorfos mediante resonancia nuclear magnética, la medición espectroscópica de iones de hidrógeno y la física del positronium. MacCallum se halla en los Laboratorios Sandia. Se licenció por la Universidad de Princeton y obtuvo el doctorado por la de Nuevo México.

GLYN DANIEL ("Monumentos megalíticos") es profesor de la cátedra Disney de arqueología y director del departamento de arqueología de la Universidad de Cambridge. Nacido en el sur de Gales, comenzó su formación en Cambridge en 1932 y allí ha permanecido desde entonces. Especialmente interesado por los monumentos megalíticos y la historia de la arqueología, ha escrito una serie de libros sobre estos temas.

MANUEL GARCIA VELARDE y CHRISTIANE NORMAND ("Convección") han colaborado estrechamente sobre el tema de su artículo desde 1974 en que ambos coincidieron en la división de física teórica del Centro de Investigación Nuclear de Saclay en Francia. Velarde es catedrático de mecánica estadística; hasta hace poco (junio de 1980) dirigió el departamento de física de fluidos de la Universidad Autónoma de Madrid y, en la actualidad, tiene a su cargo el departamento de física fundamental de la Universidad Nacional de Educación a Distancia (Madrid). Doctor en ciencias físicas por la Universidad Complutense de Madrid (1968) y por la Universidad de Bruselas (1970), ha trabajado durante largos períodos de tiempo en centros de investigación de los Estados Unidos, Francia, Noruega, Bélgica e Inglaterra. Normand continúa trabajando en Saclay, donde empezó como becaria en 1973, tras su licenciatura en física (con diploma en física de plasmas) por la Universidad de París-Sur (Orsay). Desde 1976 es colaboradora de investigación del Centro Superior de Investigaciones Científicas de Francia (CNRS).

ROBERT DEGABRIELE ("Fisiología del koala") enseña biología en el Riverina College of Advanced Education, que se encuentra en la ciudad interior de Wagga Wagga, en Nueva Gales del Sur, Australia. Estudió en la Universidad de Nueva Gales del Sur, doctorándose en biología en 1977. Su interés por los marsupiales y su evolución, escribe, "surgió como resultado de ser alumno de T. J. Dawson, autor del artículo "Canguros" en el número de octubre de 1977 de *Investigación y Ciencia*".

WILLIAM R. BAUER, F.H.C. CRICK y JAMES H. WHITE ("ADN superenrollado") abordan el problema del superenrollamiento desde bases muy distintas. Bauer es profesor de microbiología de la Universidad estatal de Nueva York en Stony Brook. Tiene dos licenciaturas, en química, por el Instituto de Tecnología de California, y en fisiología animal y bioquímica, por la Universidad de Oxford. Se doctoró en química por el Instituto de Tecnología de California. Antes de trasladarse a Stony Brook, en 1973, enseñó química durante varios años en la Universidad de Colorado en Boulder. Crick, ganador en 1962 del premio Nobel de medicina y fisiología, junto con James D. Watson y M.H.F. Wilkins, por el descubrimiento de la estructura en doble hélice del ADN, es profesor investigador de la cátedra Kieckhefer del Instituto Salk para Estudios Biológicos. Hizo la carrera de física en el University College de Londres y obtuvo el doctorado en biología por la Universidad de Cambridge en 1954. White enseña matemática en la Universidad de California en Los Angeles. Tras graduarse por la Universidad de California en Berkeley, obtuvo su doctorado en matemática por la Universidad de Minnesota en 1968. Es especialista en topología de curvas y superficies en el espacio tridimensional.

E. EUGENE LARRABEE ("La hélice") es profesor asociado de aeronáutica y astronáutica del Instituto de Tecnología de Massachusetts. Entró en el claustro de profesores del MIT como profesor ayudante en 1946, según dice él mismo, "después de trabajar en la guerra como técnico aerodinámico en la Curtiss-Wright Corporation. Siempre he estado interesado por los aviones, habiendo quedado profundamente impresionado cuando niño por el ejemplo de Charles Lindbergh. Mis afanes profesionales se centran en el automóvil y el avión, y he trabajado en materias que van desde la estrategia del control del conductor hasta el control lateral de aviones para vuelo a vela."

Hace...

José M.ª López Piñero

... cuatrocientos cincuenta años

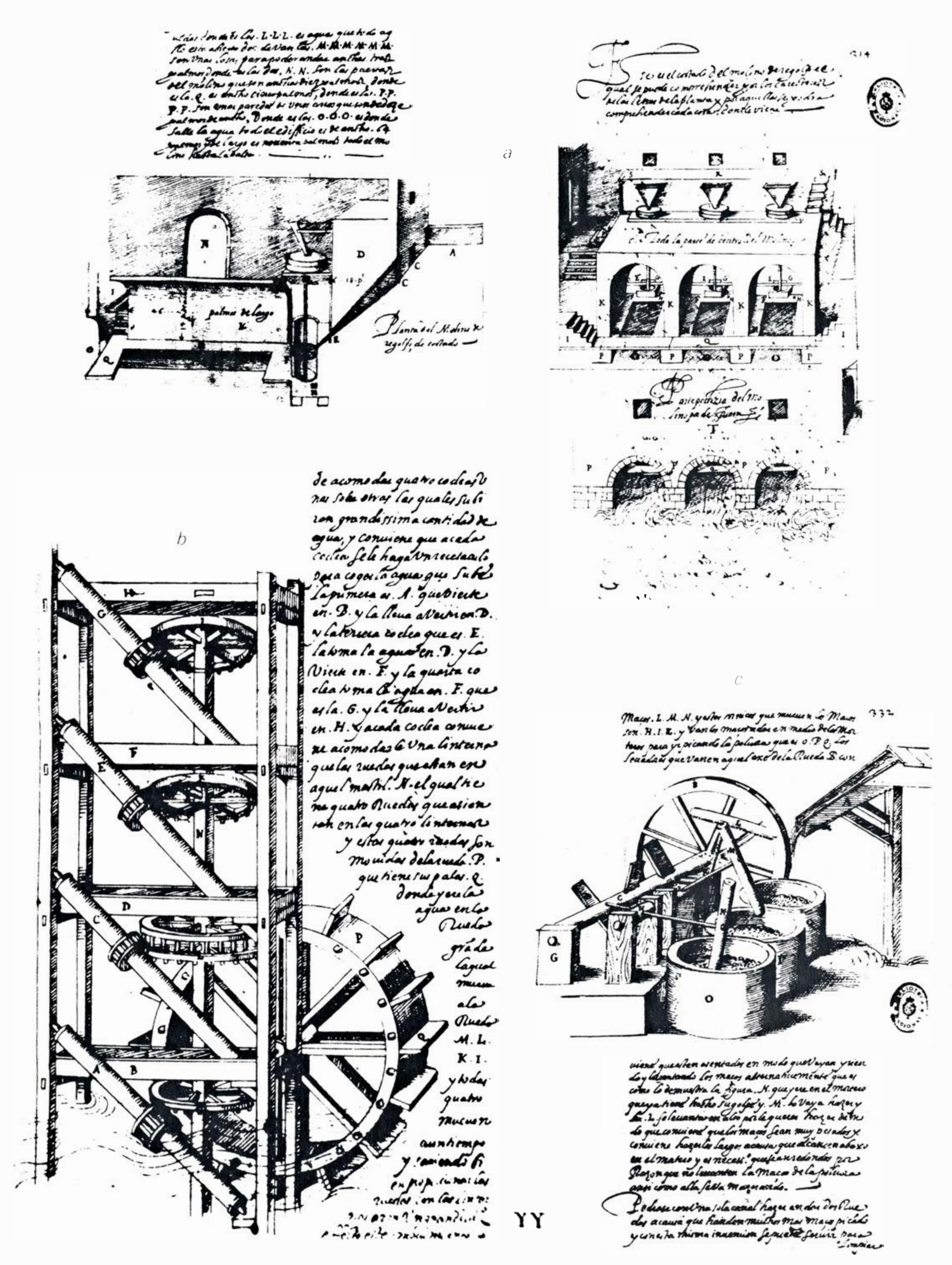
Juanelo Turriano entró en relación con Carlos I de España, cuando éste fue coronado emperador en Bolonia. Regalaron entonces al monarca el famoso *astrarium* o reloj astronómico que Giovanni de Dondi había construido en 1364. El aparato estaba estropeado y varios mecánicos no consiguieron repararlo. Turriano, consultado también con este motivo, opinó que era una tarea irrealizable pero ofreció construir otro nuevo. Aceptado el ofrecimiento por Carlos I, "tardó, como él me ha dicho –afirma el cronista Ambrosio de Morales–, en imaginarlo veinte años enteros, y de la gran vehemencia y embebecimiento del considerar, enfermó dos veces", mientras que la construcción la terminó en tres años. "Tiene el reloj todo –sigue diciendo Morales– mil y ochocientas ruedas sin otras muchas cosas de hierro y latón... Preguntóle el Emperador qué había de escribir en el reloj. El respondió que esto: *Ianelius Turrianus Cremonensis horologiorum architector.* Parando él aquí, añadió su Majestad: *facile princeps.*" Esta anécdota recogida por el cronista refleja el gran aprecio que Carlos I tuvo por Turriano, que continuó a su servicio hasta la muerte del monarca en Yuste. Le acompañó a su retiro en dicho monasterio, donde el monarca acudía diariamente a su taller para ayudarle en la construcción de sus máquinas e instrumentos. En 1558, Turriano continuó al servicio de Felipe II, pasando de su primitivo cargo de "relogero" a ostentar el título de "matemático mayor".

Turriano desarrolló una sobresaliente actividad en relación con muy diversos "ingenios" y "artificios". Construyó relojes y autómatas, uno de los cuales parece que se conserva en el Kunsthistorisches Museum de Viena. Este aspecto de su obra impresionó vivamente la imaginación popular y durante mucho tiempo la tradición habló de su "hombre de palo", autómata que, según Ponz, desde su casa en Toledo "iba a la del arzobispo donde tomaba la ración de pan y carne, haciendo varias cortesías al ir y al volver". Inventó también varias máquinas, entre ellas "un molino de hierro tan pequeño que se puede llevar en la manga y muele más de dos celemines de trigo al día, moviéndose él a sí mismo". En su diseño y construcción participó personalmente Carlos I durante sus últimos días en Yuste, ilusionándole al emperador la importancia que tendría su aplicación al abastecimiento de las tropas en plena campaña. En la conocida obra de Ramelli *Le diverse et artificiose machine* (1588) figura un molino de este tipo, que probablemente es el de Turriano, o al menos inspirado en él.

El aspecto más importante de la actividad de Turriano fue, sin embargo, la ingeniería hidráulica. Participó en empresas como el canal de Colmenar o el gran pantano de Tibi y, sobre todo, construyó el "artificio" para elevar el agua del Tajo a la ciudad de Toledo, sin duda la más famosa obra de ingeniería de la centuria. Hizo previamente "su modelo en pequeñita forma" y acabó la construcción de un primer "artificio" a mediados de 1569. Cinco años más tarde firmó un nuevo convenio para elevar mayor caudal de agua, terminando un segundo "artificio" en 1581. Morales describe la máquina en los siguientes términos: "La suma de ella es engoznar unos maderos pequeños en cruz por en medio y por los extremos... estando todo así encadenado, al moverse los dos primeros maderos junto al río se mueven todos los demás". La energía necesaria procedía de una rueda de paletas movida por la corriente del río. Se calcula que fue capaz de elevar durante casi un siglo 16.000 litros de agua diarios desde una distancia de 600 metros a una altura de casi cien.

Poco antes de iniciar sus trabajos para el "artificio" toledano, Turriano terminó, al parecer, el tratado que lleva el título de *Los Veinte y Un Libros de los Ingenios y Máquinas de Juanelo,* desde que Juan Gómez de Mora, arquitecto mayor de Felipe III y Felipe IV, ordenó el manuscrito a mediados del siglo XVII. Se trata de cinco volúmenes en folio, con más de novecientas páginas, que incluyen muchos centenares de dibujos de elevada calidad técnica y artística y sus explicaciones en un castellano entrecortado y lleno de italianismos. Su tema central es la ingeniería hidráulica, pero en torno al mismo se tratan otras muchas materias. Los dos primeros volúmenes se ocupan de las técnicas para hallar agua, de los betunes y cementos utilizados en las conducciones, de los acueductos, de las fuentes, cisternas y aljibes, etc. El tercer volumen es una completa exposición acerca de los molinos, especialmente hidráulicos, pero también de viento o movidos por animales, detallando sus aplicaciones de todo tipo: fabricación de harina, azúcar y aceite, descascarillado del arroz, preparación de la pólvora, batanes, etcétera. Contiene también este volumen uno de los escasos estudios de la época sobre técnicas industriales químicas y textiles, como la obtención de ceras y almidón, sal, nitro y alumbre, o el lavado y el tintado de las lanas. El tema del volumen cuarto es la construcción de puentes de todas clases, así como de la forma de cortar la madera y la piedra, fabricación de mortero y yeso y técnicas para manejar y transportar grandes pesos. El quinto, por último, se ocupa de la construcción y defensa de los puertos, así como de algunos problemas relacionados con el trabajo y la inmersión debajo del agua.

Reti ha subrayado que los diseños de Turriano incluyen medidas y cálculos de rendimiento, en contraste con los proyectos imaginativos e irrealizables que ocupan buena parte de la literatura técnica de la época, incluidos los célebres libros de Jacques Besson y Agostino Ramelli. Se refieren a "ingenios" y máquinas reales y no a pretendidos inventos nuevos y espectaculares, por lo que el tratado, en opinión de Reti, es "una revisión casi completa de las artes mecánicas conforme eran conocidas y practicadas en el siglo XVI... que ofrece numerosos detalles sobre operaciones técnicas acerca de las cuales no existe hoy ninguna otra información disponible". Ello no quiere decir que no incluya también aportaciones personales. Reti, que se ha interesado especialmente por las turbinas hidráulicas horizontales, destaca dos diseños suyos: el "molino de bomba", con una rueda de hojas en espiral, y el "molino de regolfo", provisto de una rueda de cinco o seis hojas helicoidales que funciona en el interior de un cilindro. Sus rendimientos eran semejantes o superiores a los mejores molinos de viento modernos. Reti afirma asimismo que "la cinemática práctica de Juanelo era más avanzada y racional que la de cualquiera de sus contemporáneos", citando expresamente a Vanoccio Biringuccio, Georgius Agricola, Jacques Besson, Agostino Ramelli y Girolamo



"Molino de regolfo" (**a**), *"máquina para picar la pólvora"* (**c**) *e "ingenio para subir al agua"* (**b**). *Diseños de* **Los Veinte y Un Libros de los Ingenios y Máquinas**

Cardano. Este último, por cierto, cita en *De subtilitate* (1550) a Turriano, como "hombre de gran ingenio en todas las cosas que se refieren a las máquinas".

El manuscrito de Turriano no fue publicado en su tiempo quizá porque muchos de los materiales que contenía, recogidos desde un puesto con el respaldo del poder real, fueron considerados confidenciales. A finales del siglo XVIII, se proyectó una edición por recomendación del matemático y físico Benito Bails, que no llegó a aparecer. Tampoco se ha publicado la traducción inglesa anotada que Reti y el profesor británico Alex Keller anunciaron hace unos años, quizá por la dedicación del primero al estudio de los códices madrileños de Leonardo da Vinci durante la época inmediatamente anterior a su muerte. En la actualidad, el análisis de la obra está siendo realizado por J. A. García-Diego que ha publicado, entre otros trabajos, un estudio sobre su "libro" acerca de las presas. En él plantea la cuestión de si el autor fue realmente Turriano. Sus dudas parten del informe de Bails y se apoyan en diversos detalles del contenido del manuscrito que también llamaron la atención de Reti antes de su fallecimiento.

Aunque Turriano era un "mecánico" sin estudios académicos, se preciaba de tener una buena preparación científica, sobre todo en cosmografía y matemática. Colaboró en las observaciones de los eclipses de 1577 organizadas por Juan López de Velasco, y se ocupó de modo particular en las cuestiones relacionadas con el calendario. No solamente adaptó su *astrarium* a los nuevos cómputos, sino que fue uno de los numerosos científicos que en España contribuyeron a los estudios preparatorios de la reforma gregoriana. Juan de Herrera tenía en su biblioteca dos manuscritos suyos sobre el tema, titulado el primero *Discurso sobre la nueva reformación del año* y el segundo redactado en italiano.

ERRATUM

En el número 47, correspondiente al mes de agosto de este año, en la página 4, línea 4 de la primera columna, donde dice "Valencia el año 1683", debe decir "Valencia el año 1638".

La traducción de los lenguajes de computador

Los traductores de los lenguajes de alto nivel son programas grandes y complejos y conviene hacerlos independientes de la máquina en que se ejecutan para evitar duplicación de esfuerzos

Manuel Alfonseca y María Luisa Tavera

La existencia de los seres vivos depende del fenómeno de la transmisión y la recepción de información. Incluso el ser unicelular más elemental precisa conocer datos sobre el ambiente que le rodea para llevar a cabo sus actividades vitales y aumentar sus posibilidades de supervivencia. Esta necesidad se multiplica enormemente cuando varios organismos elementales se agrupan para constituir un ente de orden superior, cuya viabilidad depende de la existencia de una coordinación perfecta entre los diversos individuos que lo componen. Hace unos 800 millones de años aparecieron en la Tierra los seres pluricelulares: hongos, metafitos o vegetales y metazoos o animales. Las fuerzas de la evolución dieron lugar a la aparición, en muchos de estos organismos, de sistemas de transmisión y recepción de información: los diversos sentidos permiten recibir datos sobre el ambiente exterior; los sistemas nervioso y hormonal gobiernan el reparto interno de la información a través de impulsos eléctricos del primero y de las concentraciones de ciertas sustancias químicas el segundo.

La sociedad humana puede considerarse como un superorganismo formado por multitud de entes individuales, unidos entre sí de una forma semejante, aunque menos rígida, que las células que componen el cuerpo de un ser pluricelular. Como tal organismo complejo, la sociedad precisa de mecanismos de transmisión de información para poder organizarse y sobrevivir. El más importante de estos mecanismos es el lenguaje.

Desde su aparición, hace al menos dos millones y medio de años, el hombre ha utilizado herramientas. Estas han ido complicándose con el tiempo, hasta convertirse en máquinas complejas. Hasta hace poco más de 30 años, todas estas máquinas podían gobernarse mediante unas pocas órdenes, de naturaleza principalmente mecánica, relativamente fáciles de aprender y practicar. Con la invención del computador digital, sin embargo, apareció una máquina capaz de realizar rápidamente numerosas operaciones complejas, cuyo control exige la especificación de gran cantidad de información. El problema de comunicarse con un computador es, por tanto, comparable al de la comunicación de dos seres humanos entre sí, y se ha resuelto de una forma semejante: mediante la adopción de un lenguaje común.

Obviamente, un lenguaje destinado a comunicarse con un computador no puede ser igual a los que utilizamos los seres humanos para entendernos. Los lenguajes naturales, así llamados porque se han originado y han evolucionado espontáneamente, son extremadamente complicados y contienen ambigüedades gramaticales, términos cuyo significado depende, no ya del contexto, sino de toda la masa de conocimientos previos del interlocutor, etcétera. Para comunicarse con una máquina es necesario darle órdenes concretas, claras y exentas de ambigüedad. Además, es preciso transmitirlas de tal forma que el computador sea capaz de recibirlas e interpretarlas. ¿Qué clase de señales transmisoras de información puede "comprender" un computador?

Los computadores electrónicos digitales de la actualidad sólo son capaces de distinguir entre dos estados de tensión eléctrica de un cable, o entre dos estados de polaridad de un material magnético. La tensión puede ser alta o baja; la imantación, norte o sur. Puesto que las variables transmisoras de la información sólo pueden adquirir dos valores distintos, se dice que la información está expresada en código binario. Para facilitar la notación, los seres humanos que tenemos que entrar en relación con el computador solemos representar los dos estados posibles con los símbolos 0 y 1. Se denomina bit a la cantidad de información que puede

0	0100 000101 000100
1	0101 000110 000100
2	0110 000101 000100
3	1000 000000 000000
4	0000000000000101
5	0000000000000010
6	0000000000000001

CONTENIDO DE LA MEMORIA DE UN COMPUTADOR HIPOTETICO. Cada posición de memoria consta de 16 bits de información. (Se llama bit a la unidad de información.) Las cuatro primeras almacenan un programa, mientras que las tres últimas contienen los datos para el mismo. Cada instrucción consta de un campo de operación (los cuatro primeros bits), que indica la acción a realizar, y dos campos de dirección de seis bits, que especifican los operandos sobre los que se va a efectuar la operación. Este programa podría realizar las siguientes acciones: sumar el contenido de la posición de memoria número 5 (000101 en código binario) al de la número 4 (000100). Restarle el contenido de la número 6 (000110). Multiplicarla por el de la número 5. Finalmente, detenerse. Si los valores iniciales son los indicados, el valor final de la posición número 4 sería 0000000000001100: el número decimal 12.

expresarse precisamente con uno solo de estos símbolos.

Los computadores disponen de una "memoria", donde pueden almacenar las órdenes que deben cumplir, y los datos sobre los que han de actuar. Esta memoria se compone de un número generalmente elevado de elementos, cada uno de los cuales puede almacenar un bit de información. En la práctica, el bit es una unidad demasiado pequeña, y se suele considerar la memoria dividida en unidades que contienen cierto número de bits, generalmente 8 o 16. Estas unidades se numeran consecutivamente a partir de cero. El número que corresponde a cada una se denomina su dirección.

Una orden dirigida a un computador digital recibe el nombre de instrucción. De acuerdo con lo que acabamos de decir, una instrucción constará de una serie de bits que especifican las tensiones eléctricas altas o bajas, o las polaridades magnéticas norte o sur, que deben asignarse a determinados terminales eléctricos o magnéticos. Representaremos simbólicamente una instrucción mediante una cadena de ceros y unos. Se dice entonces que la instrucción está escrita en el lenguaje de la máquina o, más brevemente, que es una instrucción-máquina, porque el computador puede entenderla directamente.

Una instrucción suele constar de varios campos bien especificados. Uno de ellos indica al computador qué operación se desea realizar, y se denomina código de operación. Otros, que pueden no existir en algunos casos, seleccionan los operandos con los que se va a realizar la operación. Por ejemplo: la instrucción máquina.

0100 000101 000100

puede significar que se desea obtener la suma de los datos contenidos en las posiciones de memoria indicadas por las direcciones 000101 y 000100. Por el contrario, la instrucción

0101 000101 000100

que sólo se diferencia en el código de operación, puede indicar que la operación que debe efectuarse entre los mismos datos es una resta.

Un conjunto ordenado de instrucciones máquina que el computador debe ejecutar constituye un programa escrito en el lenguaje de la máquina. Los primeros computadores digitales, desarrollados durante la década de 1940, sólo aceptaban esta clase de programas. Pronto se comprobó, sin embargo, que programar en lenguaje de la máquina es una labor extraordinariamente tediosa y en la que el programador suele cometer numerosos errores. Además, estos programas son después muy difíciles de interpretar y analizar. El lenguaje de la máquina podrá ser la forma natural de comunicarse con un computador desde el punto de vista de éste, pero la mentalidad humana dista mucho de estar adaptada al código binario. Por consiguiente, pronto se comenzaron a buscar alternativas que permitieran escribir programas en un lenguaje más fácil de comprender.

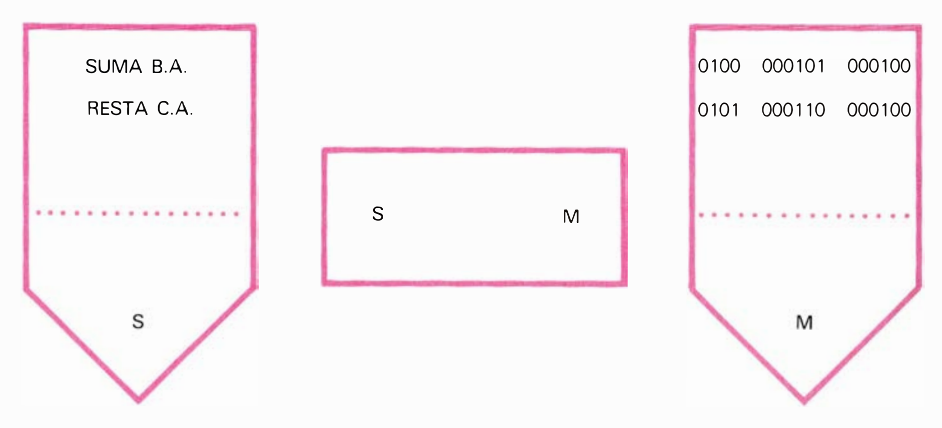
El primer paso consistió en asignar nombres simbólicos a los códigos de operación y las direcciones de las posiciones de memoria mencionadas en las instrucciones máquina, así como en la utilización directa de números decimales en lugar de la notación binaria. Si se aplica esta codificación a una instrucción máquina se obtiene otra cuyo significado es considerablemente más claro a nuestros ojos que el de una simple cadena de ceros y unos. Por ejemplo, las dos instrucciones citadas más arriba podrían expresarse así:

SUMA *B*, *A*
RESTA *B*, *A*

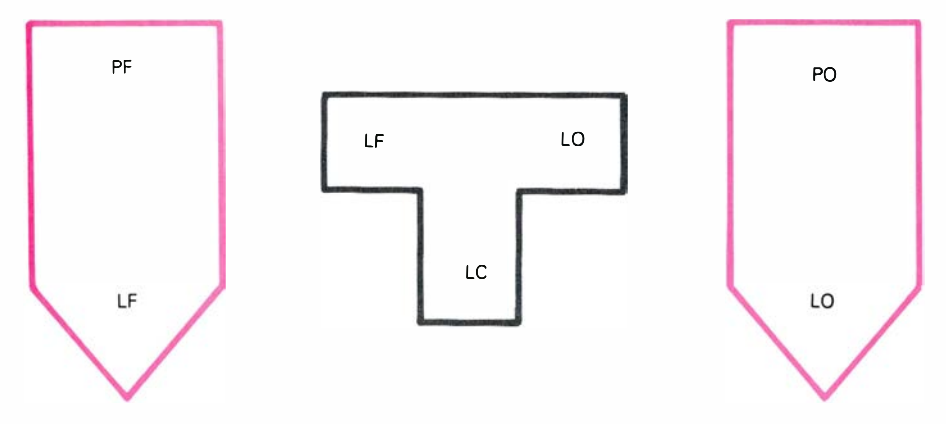
que indican que se desea sumar (o restar) el valor contenido en la posición de memoria llamada *B* al correspondiente a la posición *A*, dejando el resultado en *A*. Estos lenguajes se llaman simbólicos, y permiten liberar al programador de la tarea de recordar las direcciones concretas de memoria donde se almacenan los datos y las instrucciones de su programa. Es fácil, además, construir programas cuya ejecución no dependa de las direcciones concretas de memoria en que se almacenen. Aunque el lenguaje simbólico presente muchas ventajas desde el punto de vista humano sobre el lenguaje de la máquina, resulta evidente que un computador no puede "entender" directamente los programas simbólicos. Por consiguiente, es necesario traducirlos antes de que puedan ser ejecutados correctamente. La traducción es un proceso puramente mecánico y, por tanto, sería absurdo realizarlo a mano, puesto que disponemos de un computador cuya utilidad principal es precisamente la de liberar al hombre de la realización de tareas rutinarias. Por tanto, es lógico construir un programa cuya acción consista en aceptar programas simbólicos y traducirlos a otros equivalentes escritos en el lenguaje de la máquina. Este traductor se llama ensamblador, y aunque parece a primera vista que en su construcción es ineludible utilizar el lenguaje de la máquina, existen diversas técnicas, en cuya descripción no vamos a entrar aquí, que permiten reducir este esfuerzo al mínimo indispensable.

	SUMA	*B, A*
	RESTA	*C, A*
	MULT	*B, A*
	ALTO	
A	CONS	3
B	CONS	2
C	CONS	

PROGRAMA EN LENGUAJE SIMBOLICO equivalente al programa en lenguaje de máquina de la figura anterior. Los códigos de operación han sido sustituidos por códigos simbólicos (SUMA, RESTA, MULT, ALTO) que indican con mayor claridad la operación que se desea efectuar. A ciertas direcciones de memoria se les asignan nombres mnemotécnicos o etiquetas. El programador que utiliza el lenguaje simbólico no tiene que preocuparse de recordar las direcciones concretas donde van a introducirse sus datos en la memoria del ordenador. Podemos construir un programa que se ejecute correctamente con independencia de las posiciones de memoria donde se almacene.



SE DENOMINA ENSAMBLADOR a un programa que traduce programas escritos en lenguaje simbólico (S), dando lugar a otros equivalentes escritos en el lenguaje de la máquina (*M*). Existen técnicas que permiten escribir el ensamblador en el propio lenguaje simbólico que se va a traducir. Los programas se representan aquí mediante cajas picudas y los ensambladores, como rectángulos.



PROCESO DE COMPILACION de un lenguaje de alto nivel. Los programas escritos en un lenguaje de alto nivel no son directamente ejecutables en un computador. El programa que se desea traducir se denomina programa fuente (abreviado *PF*). El resultado de la traducción constituye el programa objeto (*PO*). El compilador (representado en el diagrama superior por la figura en forma de *T*) acepta cualquier programa escrito en el lenguaje fuente (*LF*) y genera otro programa equivalente escrito en el lenguaje objeto (*LO*). *LC* es el lenguaje del compilador. No es necesario que *IC* sea el mismo lenguaje que *LO* o *LF*.

Puesto que a cada instrucción máquina se le hace corresponder una instrucción simbólica, un programa escrito en lenguaje simbólico tendría, en principio, el mismo número de instrucciones que su contrapartida escrita en el lenguaje de la máquina, obtenida como resultado de la traducción realizada por un ensamblador. Sin embargo, en la práctica, los lenguajes simbólicos incorporan ciertas instrucciones especiales, llamadas seudoinstrucciones, que no generan código ejecutable alguno, y cuya misión es transmitir al ensamblador información respecto a cómo debe llevar a cabo su función.

Pero los lenguajes simbólicos no constituyen una solución definitiva al problema de facilitar la comunicación con el computador. Para utilizarlos, es preciso conocer bastante a fondo la estructura física de la máquina en la que se va a ejecutar el programa. Esto significa, por una parte, la reducción del número de usuarios potenciales del computador, y, por otra, la imposibilidad de construir programas independientes de la máquina y, por ende, una enorme duplicación de esfuerzos. Como consecuencia de estas necesidades se desarrollaron los lenguajes de alto nivel [véase "Lenguajes de programación", por Jerome A. Feldman, INVESTIGACIÓN Y CIENCIA, febrero de 1980]. Estos proporcionan al programador la posibilidad de utilizar lenguajes más próximos a los naturales, haciendo posible la construcción de programas más concisos, más fáciles de entender y corregir, así como potencialmente independientes de la máquina concreta sobre la que han de ser finalmente ejecutados.

Naturalmente, los programas escritos en un lenguaje de alto nivel tampoco son directamente ejecutables en un computador. Antes de poder hacer uso de ellos es necesario disponer de traductores que los conviertan en otros que la máquina pueda comprender. Existen dos clases principales de estos traductores: los compiladores y los intérpretes.

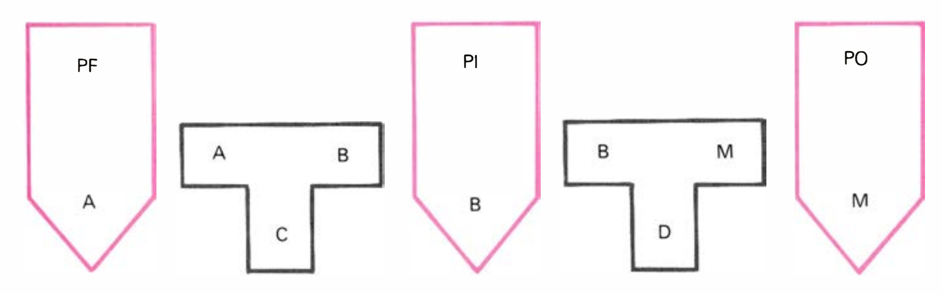
Un compilador es un traductor que acepta programas escritos en un lenguaje de alto nivel y genera programas equivalentes escritos en un lenguaje diferente. Obsérvese que no especificamos que este segundo lenguaje tenga que ser necesariamente el de la máquina. En el proceso de la compilación pueden distinguirse los siguientes elementos: *a*) el programa que el compilador acepta para su traducción, que se denomina programa fuente; *b*) el lenguaje fuente, en el que está escrito el programa fuente; *c*) el lenguaje objeto, al que se desea traducir el programa fuente; *d*) el programa objeto, resultado de la traducción; *e*) el lenguaje en que está escrito el propio compilador, que no tiene por qué coincidir con los lenguajes fuente y objeto.

Evidentemente, si se desea obtener un programa directamente ejecutable sobre un computador determinado, debemos llegar más pronto o más tarde a la generación de un programa escrito en el lenguaje de la máquina de que se trate. Sin embargo, no es necesario realizar este proceso de una sola vez. Por ejemplo, si el programa que queremos traducir está escrito en el lenguaje *A*, y debe ejecutarse sobre la máquina *M*, puede ocurrir que no dispongamos de un compilador que realice directamente la traducción, pero que sí estén a nuestro alcance los dos traductores siguientes: un compilador del lenguaje *A* al lenguaje *B* y un compilador del lenguaje *B* al de la máquina *M*. En este caso, podemos realizar la traducción en dos etapas, en la forma que indica la ilustración inferior de esta página.

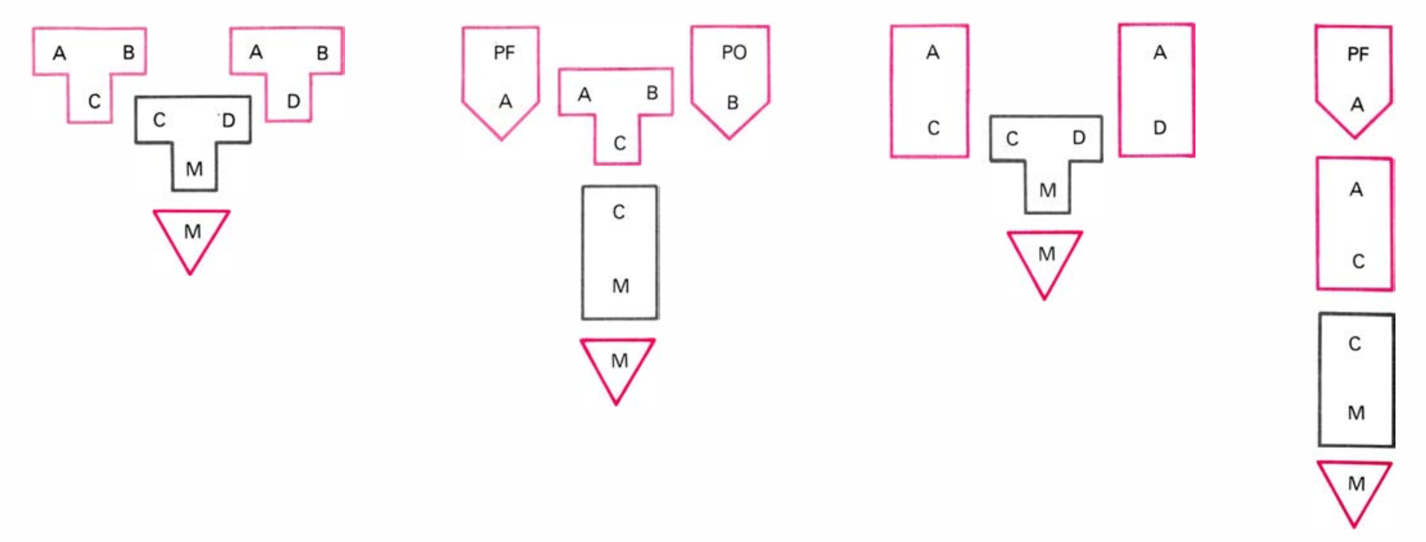
Un intérprete es un traductor que acepta programas escritos en un lenguaje de alto nivel y ejecuta cada una de sus instrucciones a medida que las encuentra. Se diferencia de un compilador en que no genera código traducido. Por consiguiente, no existen en este caso el lenguaje objeto ni el programa objeto.

Tanto los compiladores como los intérpretes son, a su vez, programas que deben ejecutarse en algún computador concreto. Es cierto que el traductor puede escribirse en cualquier lenguaje, incluso de alto nivel, pero si se desea ejecutarlo será necesario traducirlo al lenguaje de la máquina, ya sea compilándolo o interpretándolo. Todas las combinaciones son posibles, es decir, se puede: compilar un compilador, interpretar un compilador, compilar un intérprete, e interpretar un intérprete.

En la práctica, el proceso de la traducción introduce siempre pérdidas de eficacias y redundancias. En extremadamente difícil construir un compilador que genere código cuya eficiencia sea comparable a la de un programa equivalente, escrito directamente en el lenguaje objeto por un programador experto que haga uso de todos los trucos del oficio. Por otra parte, la interpretación de un programa suele ser menos eficaz que el resultado de su



PROCESO DE COMPILACION del lenguaje fuente *A* al lenguaje de la máquina *M*, realizado a través de un paso intermedio. El primer compilador traduce el programa fuente *PF*, generando un programa intermedio *PI*, escrito en el lenguaje *B*. *PI* pasa ahora a convertirse en el programa fuente para el segundo compilador, que genera el programa objeto definitivo, escrito en el lenguaje de máquina *M*.

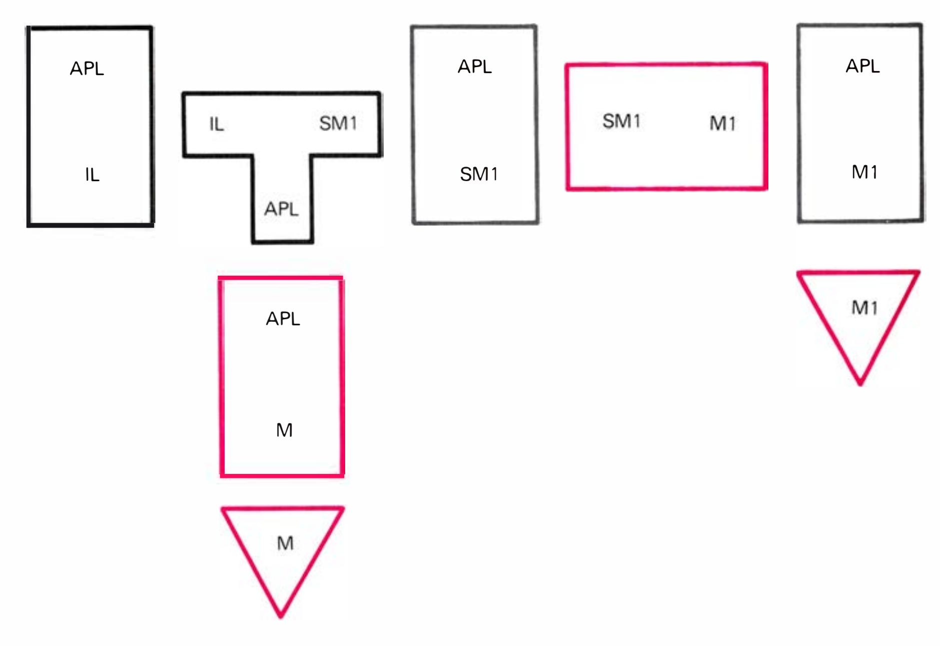


LOS TRADUCTORES SON TAMBIEN PROGRAMAS. En la primera figura empezando por la izquierda, un compilador del lenguaje *A* al *B*, escrito en el lenguaje *C*, es compilado, generándose un compilador entre los mismos lenguajes, pero escrito en el lenguaje *D*. En el segundo bloque ilustrativo, el mismo compilador es interpretado, y, por tanto, ejecutado, a través de un intérprete del lenguaje *C* para la máquina *M*. En la tercera ilustración empezando por la izquierda, un intérprete del lenguaje *A*, escrito en lenguaje *C*, es compilado, generándose otro intérprete de *A*, escrito esta vez en el lenguaje *D*. En el bloque figurativo de la derecha, el mismo intérprete es traducido y ejecutado por un intérprete del lenguaje *C* para la máquina *M*.

traducción por un compilador, puesto que el intérprete debe traducir las instrucciones fuente tantas veces como se ejecuten, mientras que el compilador genera código una sola vez por instrucción. Esto no quiere decir que un compilador sea siempre superior a un intérprete. Es fácil, por ejemplo, hacer del segundo un sistema interactivo, en el que sea mucho más sencilla y rápida la búsqueda y eliminación de los errores que inevitablemente se introducen en la construcción de cualquier programa que no sea trivial. Además, algunos lenguajes de alto nivel no son fácilmente compilables, pues el análisis y la traducción de sus instrucciones puede precisar de cierta información que no sea posible conocer hasta que el programa esté ejecutándose. En estos casos está indicada la utilización de un intérprete.

Para reducir al mínimo la pérdida de eficiencia, los compiladores y los intérpretes suelen programarse en el lenguaje simbólico asociado con el computador sobre el que se han de ejecutar. Siendo más próximo al lenguaje de la máquina, el lenguaje simbólico permite construir programas que aprovechan al máximo todas las características de aquélla, y que están prácticamente exentos de redundancia. Como contrapartida, los programas resultantes no son independientes de la máquina, puesto que cada tipo de computador suele diseñarse con su propio lenguaje máquina, lo que significa que también el lenguaje simbólico será diferente. Como consecuencia de esto, cuando aparece un computador nuevo, todos los traductores de lenguajes de alto nivel de los que se desee dotarle deben ser escritos de nuevo, sin que en general se pueda aprovechar directamente la existencia de traductores equivalentes para otros tipos de computadores.

En 1962 se introdujo el lenguaje de programación de alto nivel denominado APL (del inglés "A programming Language") destinado a usos científicos y de modelos. Se trata de uno de los lenguajes mencionados más arriba, que no se presta al proceso de la compilación. Además, ha sido diseñado con vistas a su implantación en forma de sistema interactivo, lo que significa que su traducción será, en general, tarea de un intérprete. APL es un lenguaje muy sofisticado que proporciona la posibilidad de utilizar gran número de funciones y operadores simbólicos mediante los cuales se pueden escribir programas complejos con una sintaxis muy sencilla

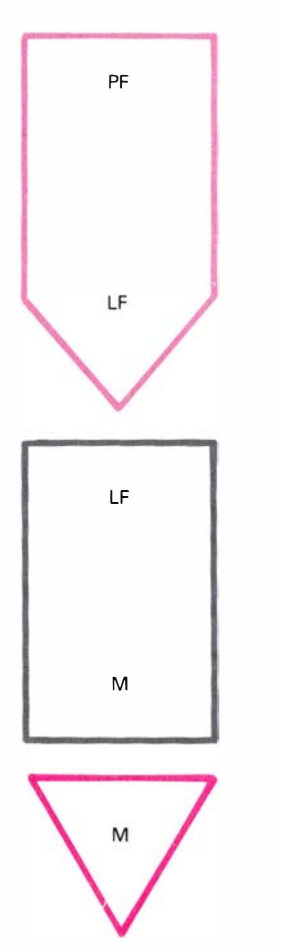


PROCEDIMIENTO UTILIZADO para generar automáticamente un intérprete del lenguaje APL que pueda ejecutarse en el computador *M*1. Se escribe primero un intérprete universal de APL en cierto lenguaje intermedio (*IL*), y también un compilador que traduzca programas escritos en *IL* al lenguaje simbólico de máquina *M*1 (*SM*1). Se compila el intérprete y se obtiene un segundo intérprete de APL escrito en *SM*1. Se ensambla entonces éste, y se genera así un tercer intérprete de APL, directamente ejecutable en la máquina *M*1. Una vez obtenido este último, tanto el compilador como los intérpretes universal e intermedio dejan de ser necesarios. Los programas que hubimos de construir se representan en negro; los que utilizamos como herramienta, pero no fueron desarrollados por los autores del artículo, aparecen en rojo; por último, se utiliza el color gris para los que fueron generados automáticamente.

```
[0]   A ← 3
[1]   B ← 2
[2]   C ← 1
[3]
[4]   S:A ← B × A + B − C
```

PROGRAMA en lenguaje intermedio (*IL*) equivalente a los de las ilustraciones de las páginas 8, 9 (*arriba*) y 10 (*arriba*). *IL* fue diseñado con una sintaxis próxima a la de los lenguajes de alto nivel, mientras que su semántica se hizo semejante a la de lenguajes simbólicos. De este modo se ha conseguido que los programas escritos en lenguaje intermedio (*IL*) sean independientes de la máquina, mientras que resulta muy sencillo construir compiladores que traduzcan el lenguaje intermedio (*IL*) a diversos lenguajes simbólicos.

y en forma altamente concisa. Las funciones y operadores primitivos del lenguaje APL actúan tanto sobre escalares como sobre vectores, matrices y tensores de orden superior. Esto hace posible escribir programas relativamente exentos de bucles de cálculo,



UN INTERPRETE es un traductor que acepta programas (*PF*) escritos en un lenguaje fuente (*LF*) y los ejecuta directamente. En este caso no se genera ningún código y no existen el lenguaje ni el programa objeto. El intérprete, indicado mediante un rectángulo vertical, estará escrito en algún lenguaje de máquina (simbolizado por *M*), directamente reconocible por el computador, representado por el triángulo invertido (*abajo*).

contrarrestándose de esta forma la pérdida de eficiencia, en cuanto al tiempo de traducción, inherente al proceso de la interpretación. Además, el lenguaje contiene como funciones primitivas muchas de las operaciones comunes relacionadas con el cálculo matricial y tensorial, tales como producto matricial, producto booleano, inversión matricial, transposición, etcétera.

Los sistemas que interpretan el lenguaje APL incluyen asimismo un editor que permite al usuario construir sus propios programas, que en APL reciben el nombre de funciones definidas. Se pone a disposición del usuario un espacio de trabajo donde puede definir y ejecutar sus funciones y almacenar los datos que éstas precisan. El sistema proporciona, además, un lenguaje de órdenes que permite guardar estos espacios de trabajo en una biblioteca situada en la memoria accesoria del computador (generalmente en un disco magnético), y recuperarlos a voluntad.

Como consecuencia de todo esto, un intérprete de APL es un programa relativamente grande y complicado, que suele constar de varias decenas de miles de instrucciones en lenguaje simbólico. Por consiguiente, su construcción requiere un esfuerzo considerable que puede cifrarse en tres o cuatro años-persona. En el Centro de Investigación de IBM de Madrid, donde trabajan los autores, se consideró la conveniencia de reducir este esfuerzo a través del diseño de un intérprete universal de APL, que pudiera adaptarse con facilidad a diversos computadores.

Para que la aplicación de nuestro procedimiento resultara justificada, habían de cumplirse las siguientes condiciones: *1*) que el esfuerzo requerido para completar un intérprete de APL, ejecutable sobre un computador que no disponga de este lenguaje, sea sensiblemente inferior a tres años-persona, y *2*) que el proceso de construcción de nuestros intérpretes no introduzca una degradación notoria en la velocidad de ejecución de éstos.

El procedimiento que hemos utilizado para la construcción de un intérprete de APL, independiente del computador, y para la obtención a partir del mismo de intérpretes ejecutables en una máquina concreta ($M1$) consta de las siguientes fases. Se construye un intérprete de APL escrito en un lenguaje intermedio (IL); se construye un compilador que traduce los programas escritos en IL, al lenguaje simbólico de la máquina $M1$; y, por último, se compila el intérprete. El producto final es un intérprete de APL escrito en el lenguaje simbólico de $M1$, que puede ser traducido ya por el ensamblador asociado a esta máquina. Si ahora se desea aplicar el mismo procedimiento para la obtención de un intérprete de APL ejecutable en una máquina diferente $M2$, bastará repetir los pasos segundo y tercero, sustituyendo $M1$ por $M2$.

Para valorar la eficiencia de nuestro procedimiento, supongamos que sea necesario generar intérpretes de APL para N máquinas distintas. Si, como se ha hecho hasta ahora, construimos directamente N intérpretes en el lenguaje simbólico de cada una de ellas, el coste total aproximado vendrá a ser $C1 = N \times W1$, donde $W1$ es el coste de cada intérprete. Por el contrario, si utilizamos nuestro procedimiento, será preciso construir un solo intérprete (escrito en lenguaje intermedio) y N compiladores, por lo que el coste total será $C2 = W1 + N \times W2$, donde $W2$ es el coste de cada compilador.

Es evidente que nuestro procedimiento estará justificado si $C2$ es significativamente inferior a $C1$, lo que se cumplirá si $W2 << W1$, es decir, si es mucho más fácil construir un compilador del lenguaje intermedio que un intérprete de APL.

Antes de poder estimar el valor $W2$ es necesario resolver dos cuestiones adicionales: selección del lenguaje intermedio y selección del lenguaje en que van a escribirse los compiladores del lenguaje intermedio. Para la solución de ambos problemas se tuvieron en cuenta todas las consideraciones anteriormente detalladas.

La cuestión del lenguaje intermedio (IL) resultó ser particularmente importante. Entre los lenguajes existentes ninguno se adaptaba completamente a nuestros propósitos, por lo que decidimos diseñar nuestro propio lenguaje de trabajo, de acuerdo con las siguientes consideraciones: *a*) debía tener una sintaxis próxima a la de los lenguajes de alto nivel. De esta forma se obtendría independencia de la máquina, objeto de nuestro trabajo, y se facilitaría la programación y búsqueda de errores, reduciendo así el valor de $W1$ en el cálculo de $C2$; *b*) debía ser semánticamente muy próximo a los lenguajes simbólicos, con objeto de que los compiladores del lenguaje intermedio fueran fáciles de construir, reduciendo de este modo el valor de

$W2$, pero que al mismo tiempo generaran código muy eficaz, sin necesidad de incluir optimizadores sofisticados. De esta forma se evitaría que la eficiencia de los intérpretes generados resultara afectada adversamente.

Para diseñar el lenguaje intermedio se siguió el siguiente procedimiento: se seleccionaron las operaciones más comunes en los lenguajes simbólicos de diversas máquinas, y se les asignó una representación próxima a la de los lenguajes de alto nivel. Además, la única suposición que se hizo respecto a los computadores, a cuyo lenguaje máquina pueda finalmente traducirse IL, consistió en suponer que su memoria estará dividida en cierto número de unidades de tamaño fijo, pero no definido, numeradas consecutivamente.

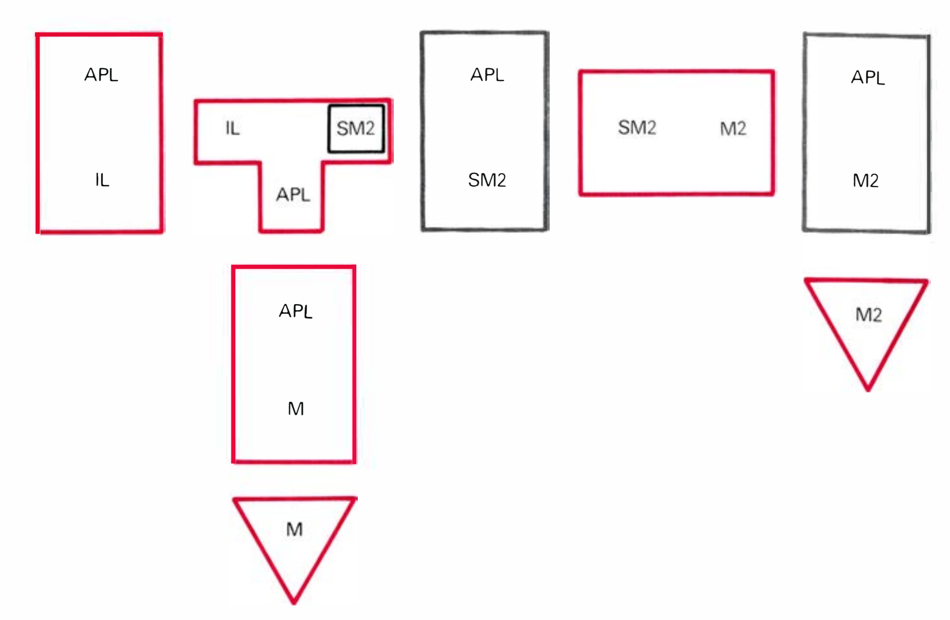
Un programa IL consta de dos clases de instrucciones: declarativas, que indican al compilador cuáles son las variables que se van a utilizar en el programa y en algunos casos les asignan valores iniciales, y ejecutables, que llevan a cabo los cálculos y acciones propiamente dichos. En el ejemplo de la ilustración superior de la página precedente, las instrucciones 0, 1 y 2 son declarativas. La número 3 separa las dos clases de instrucciones y la número 4 es ejecutable.

Para escribir los compiladores del lenguaje intermedio se eligió el lenguaje APL. De ésta forma, y gracias a su gran potencia, así como a la facilidad de programación que proporciona, se redujo enormemente el coste de construcción de cada compilador ($W2$). Es cierto que el hecho de escribirlos en este lenguaje nos obligó a someter los compiladores a una interpretación para poder ejecutarlos y realizar la traducción del intérprete universal. Como consecuencia de esto, nuestros compiladores de IL son bastante lentos. Sin embargo, debe recordarse que, en la compilación del intérprete generado, el compilador puede desecharse.

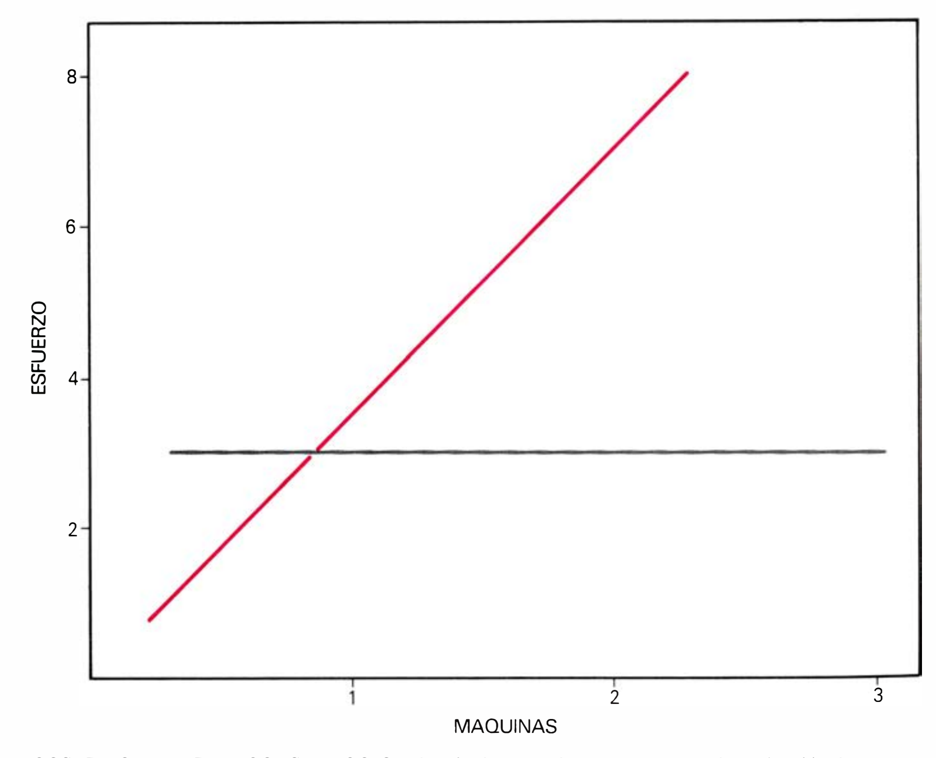
Hemos podido comprobar que el coste de fabricación de uno de nuestros compiladores es menor que dos meses persona, y, por tanto, es inferior en más de un orden de magnitud al coste de construcción de un intérprete de APL. Se verifica, por consiguiente, que $C2 << C1$, y la eficacia de nuestro procedimiento ha quedado justificada.

Durante la realización de nuestro proyecto, hemos construido los siguientes programas: un intérprete universal de APL, escrito en IL, y dos compiladores que traducen IL a los lenguajes simbólicos de dos máquinas diferentes. El intérprete universal fue traducido a los lenguajes máquina correspondientes y se comprobó que su eficiencia en cuanto a velocidad de cálculo es aceptable, en comparación con la de otros sistemas APL directamente programados en lenguaje simbólico.

Evidentemente, lo que se ha realizado para un intérprete del lenguaje APL podría haberse aplicado a cualquier traductor de cualquier otro lenguaje de alto nivel. Se ha comprobado así la utilidad de disponer de un procedimiento que permite construir traductores de lenguajes de programación que sean independientes de la máquina, y que puedan ser utilizados para reducir el esfuerzo necesario para dotar a un computador determinado de un nuevo lenguaje de programación.



AL OBJETO DE GENERAR UN SEGUNDO INTERPRETE para una máquina diferente (*M2*), sólo fue necesario rehacer la parte del compilador que genera el código objeto, indicada en el recuadro que en la ilustración aparece enmarcado en color negro. El intérprete universal y el resto del compilador son los mismos de la figura anterior, y, por tanto, pasan ahora a cumplir el papel de herramientas disponibles.



COSTE TOTAL DE LA CONSTRUCCION de *n* intérpretes de APL representado en función de *n*, para los dos métodos descritos en el texto. La recta dibujada en color rojo representa el esfuerzo necesario para escribir los *n* intérpretes directamente en los lenguajes simbólicos de las máquinas respectivas. La recta en gris corresponde a la aplicación de nuestro procedimiento. Las unidades son arbitrarias.

Lupus eritematoso sistémico

El LES, una enfermedad reumática con múltiples síntomas, es una alteración del sistema inmunitario. El paciente produce autoanticuerpos que se unen a antígenos como el ADN, formando complejos que pueden lesionar los tejidos

David Koffler

Una de las primeras manifestaciones del lupus eritematoso sistémico (LES) es una erupción cutánea en forma de mariposa que cubre la nariz y las mejillas. Este síntoma dio a la enfermedad su curioso nombre: lupus eritematosus o "lobo rojo". El nombre, no obstante, no refleja la extensión y la naturaleza polisistémica de esta enigmática enfermedad reumática que padecen una de cada 500 mujeres norteamericanas (es diez veces menos frecuente en el hombre). Se trata, desde luego, de una enfermedad diseminada, que afecta, no sólo a la piel, sino también a las articulaciones, vasos sanguíneos, corazón, pulmones, cerebro y especialmente a los riñones. Los síntomas varían en distintos pacientes y en el curso de la enfermedad.

Cualesquiera que sean los síntomas de un caso concreto, la enfermedad es siempre una alteración del sistema inmunitario. Durante algún tiempo el LES se consideró una enfermedad "autoinmunitaria" en la que el paciente producía anticuerpos que interaccionaban con antígenos de sus propios tejidos. Este puede ser el mecanismo de algunos de los síntomas de la enfermedad. No obstante, actualmente se considera que el LES se produce como consecuencia de la formación de complejos antígeno-anticuerpo que causan la lesión de los tejidos. Se desconoce si los antígenos son del propio paciente o extraños al mismo.

La función primaria del sistema inmunitario consiste en diferenciar lo propio de lo extraño y proteger el medio interno del organismo frente a infecciones y otras agresiones del medio ambiente. En el hombre, el sistema consta de alrededor de un billón (10^{12}) de células denominadas linfocitos y unas 10^{20} moléculas de inmunoglobulina o anticuerpos. Hay dos clases de linfocitos: las células *B* y las células *T*. Cada clon, o línea, de células *B* está genéticamente programada para sintetizar un determinado tipo de anticuerpo, cuyos puntos de unión reconocen una región específica de la molécula del antígeno con el que interaccionan.

En estado de reposo, un linfocito *B* presenta en su superficie, a modo de receptores, algunos de sus anticuerpos. El contacto de un receptor con su antígeno específico estimula la proliferación y diferenciación de la célula *B*. Esta forma un clon de células plasmáticas productoras de anticuerpos, que sintetizarán millones de moléculas del mismo anticuerpo. De los cinco tipos principales de anticuerpos, dos son de importancia primordial en el LES y otras enfermedades reumáticas: las inmunoglobulinas o globulinas gamma IgG e IgM. Anticuerpo y antígeno se combinan en la sangre o en los tejidos y forman un complejo inmunitario que pone en marcha varios procesos que pueden destruir las células portadoras del antígeno o dañar a espectadores inocentes: los propios tejidos del organismo.

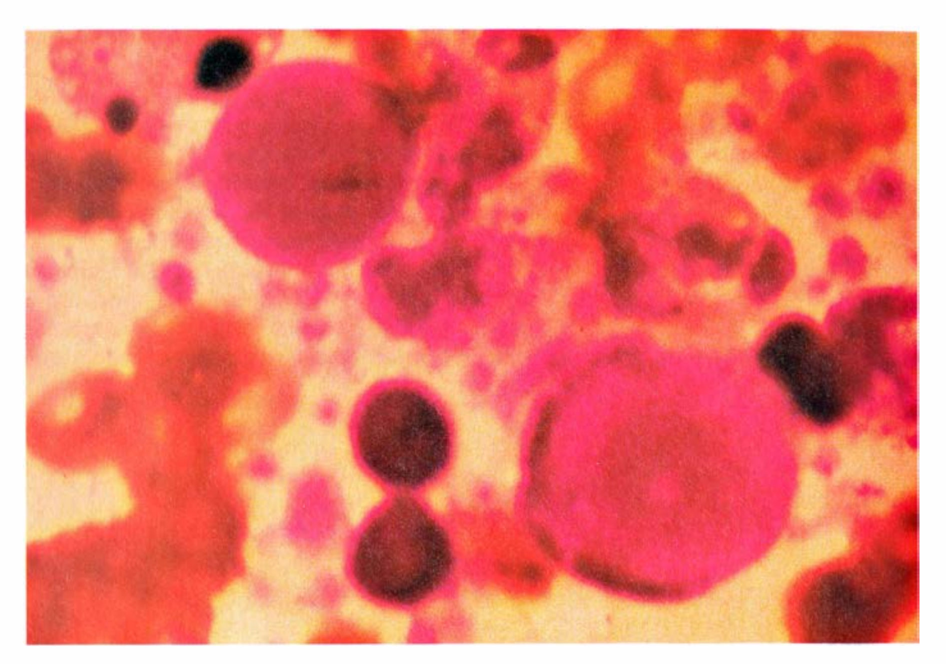
Son especialmente importantes los complejos antígeno-anticuerpo que fijan el complemento, un grupo constituido por nueve proteínas plasmáticas. Una porción del primer componente designada C1q interacciona con el anticuerpo iniciando una serie de reacciones determinantes de una respuesta inflamatoria. Algunos factores formados por el complemento se unen a los neutrófilos (los glóbulos blancos más comunes) facilitando la fagocitosis (ingestión de partículas) y la liberación de enzimas que degradan proteínas. Pequeñas moléculas derivadas del complemento atraen otras células fagocíticas. Algunos factores aumentan la permeabilidad de los vasos sanguíneos al provocar la liberación de sustancias como la histamina a partir de ciertas células; se incrementa así el aporte de células inflamatorias y el depósito de complejos inmunitarios.

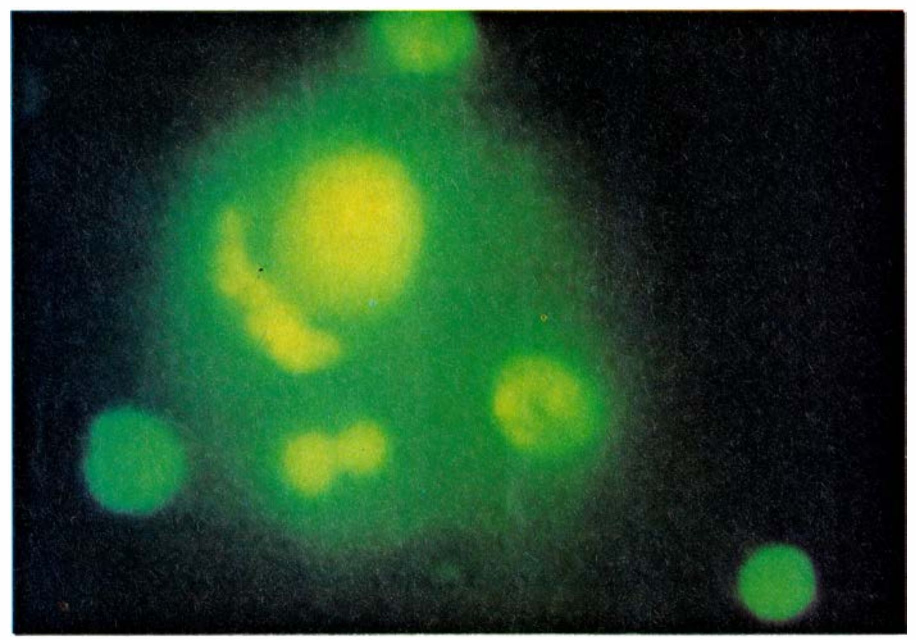
Mientras los linfocitos del tipo *B* son mediadores de la inmunidad humoral, pues están implicados en mecanismos en los que intervienen anticuerpos solubles, los linfocitos *T* median la inmunidad celular. Estos linfocitos poseen también receptores específicos en su superficie que reconocen regiones particulares de las moléculas de los antígenos. La interacción con un antígeno determina la liberación, a partir de las células *T*, de linfoquinas, sustancias químicas con múltiples efectos: producen lesiones celulares, inhiben la replicación de los virus y atraen células inflamatorias. Otros linfocitos *T* influyen en la producción de anticuerpos por las células *B*. Un tipo, las células coadyuvantes, facilita la proliferación de células *B*; otro tipo, las células supresoras, la retarda. En el LES, las células *T* parecen ser importantes debido a que sus funciones reguladoras están alteradas, y no por el hecho de que dañen directamente a las células.

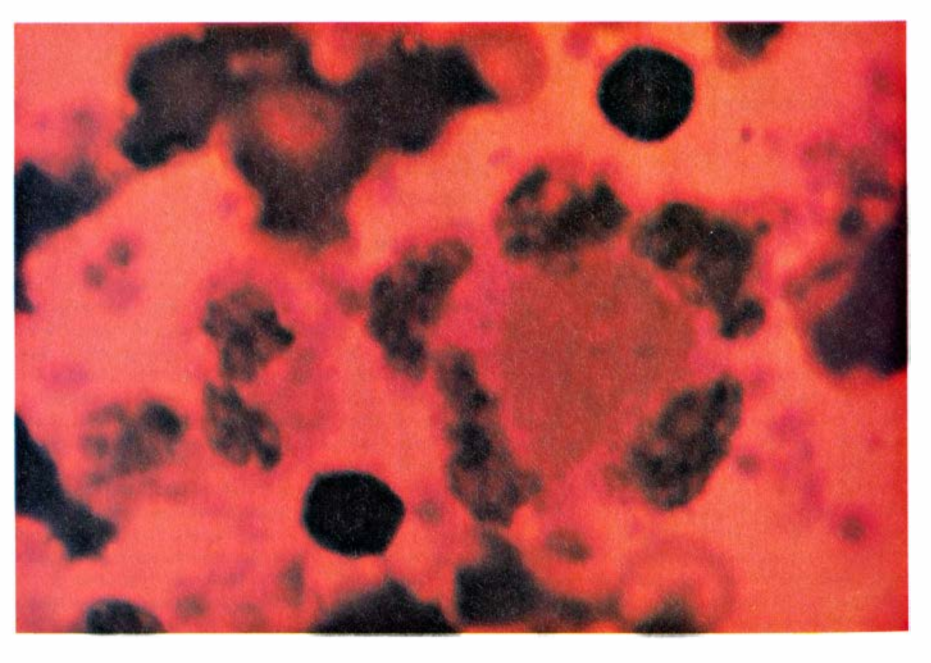
Aunque el sistema inmunitario distingue en general lo propio de lo extraño, tal distinción no es absoluta. Ya en el año 1902, Paul Ehrlich llamó la atención sobre la capacidad del sistema inmunitario de producir lesiones en el propio organismo. Ehrlich supuso que un *horror autotoxicus* protegería frente a tal posibilidad: "el organismo tiene mecanismos que evitan que la respuesta inmunitaria actúe sobre los propios tejidos". Sin embargo, esto no es cierto y en muchos casos se produce tal respuesta. Quizá la primera enfermedad autoinmunitaria producida por los médicos fuera una encefalomielitis (inflamación del cerebro y médula espinal) provocada inadvertidamente en personas que habían sido mordidas por un perro y que habían recibido múltiples inyecciones de la vacuna Pasteur para prevenir la rabia. La vacuna, preparada mediante cultivo de virus de la rabia en la médula

espinal de conejo, contenía antígenos que presentaban la misma reacción que los antígenos del sistema nervioso central humano. Es decir, el conejo y el hombre tienen en común al menos algunos de los determinantes antigénicos, por lo que los anticuerpos estimulados por la vacuna reconocían e interaccionaban con los sitios antigénicos humanos.

En 1933, Thomas M. Rivers, del Instituto Rockefeller de Investigación Médica, demostró que los monos que habían recibido una serie de inyecciones de homogeneizado de tejido nervioso desarrollaban una encefalomielitis similar. Sin embargo, hasta el año 1956 no se describió claramente la enfermedad autoinmune. En este año, Ernest Witebsky y Noel R. Rose, de la Facultad de Medicina de la Universidad de Buffalo, definieron la enfermedad autoinmune, y establecieron un método reproducible a través del cual podía anularse la tolerancia natural del sistema inmunitario hacia los antígenos propios. Estos investigadores inmunizaron conejos con extractos de glándula tiroidea de conejo mezclada con un adyuvante (una sustancia que aumenta la respuesta inmunitaria); los conejos desarrollaron tiroiditis, una inflamación de la glándula tiroides. Witebsky y Rose definieron como criterio de enfermedad autoinmunitaria lo siguiente: un antígeno tal como una proteína tiroidea inicia una respuesta inmunitaria acompañada de una serie de cambios patológicos limitados al órgano que contiene el antígeno. Ulteriormente, William O. Weigle, de la Clínica y Fundación de Investigación Scripps en La Jolla, California, observó una tiroiditis similar producida al inmunizar conejos con proteínas tiroideas de otras especies animales. Las proteínas eran reconocidas como extrañas, pero tenían deter-



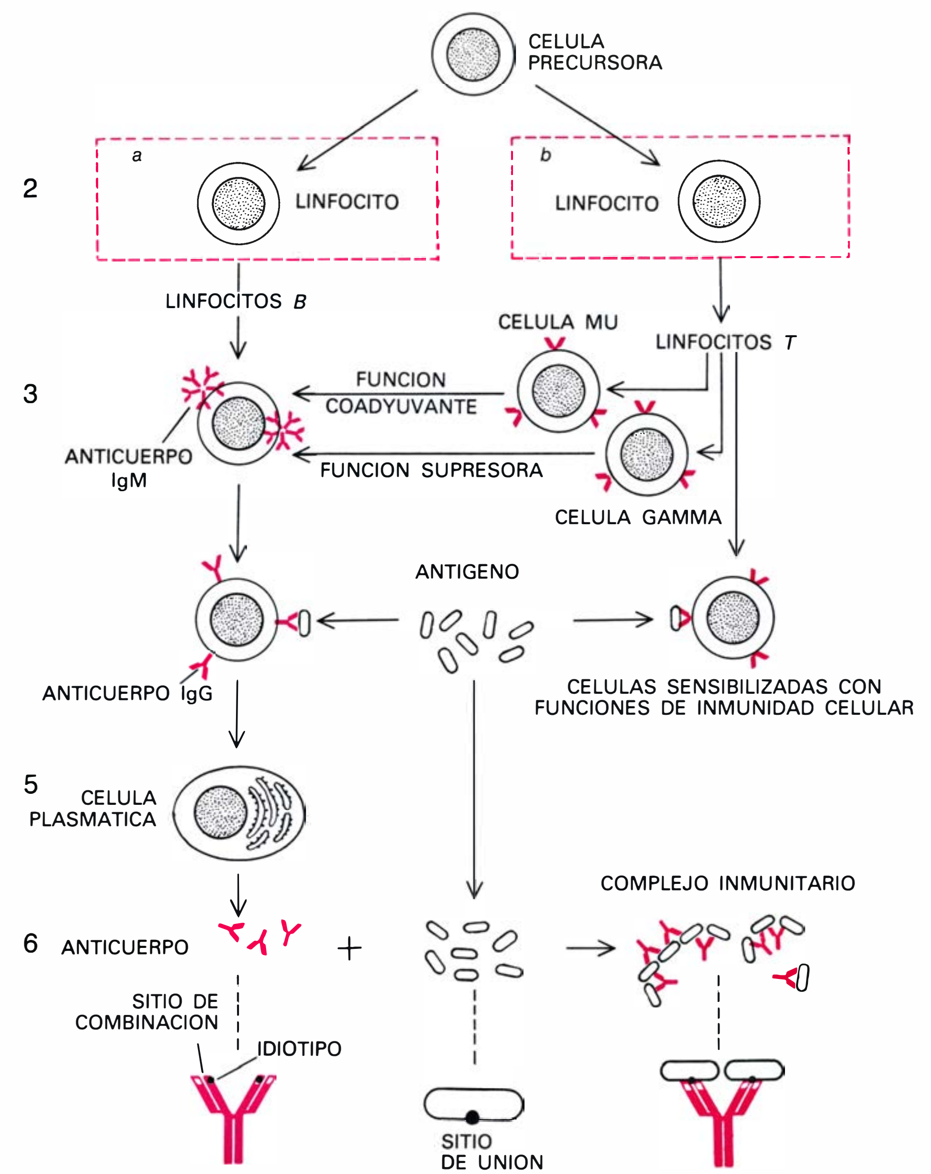




AUTOANTICUERPOS que inducen la formación de una célula del lupus eritematoso (LE). Este fenómeno se utiliza como prueba para diagnosticar la enfermedad. Los glóbulos blancos de la sangre se agitan en un tubo de ensayo para lesionarlos y se incuban a continuación con el suero del paciente. Si el suero contiene anticuerpos antinucleares característicos del LES, que reaccionan con el complejo ADN-proteínas del núcleo, tales anticuerpos penetrarán en el interior de las células alteradas por la agitación y determinarán un aumento del volumen nuclear seguido de la degeneración de los núcleos y de la liberación de cuerpos nucleares (*arriba a la izquierda, en la micrografía superior*). Un neutrófilo puede englobar el cuerpo nuclear y formar una célula LE (*abajo a la derecha, en la micrografía superior*). El núcleo del neutrófilo queda comprimido y desplazado a la periferia de la célula, como se aprecia más claramente tiñendo el ADN (*micrografía central*). Algunas veces, un cuerpo nuclear queda rodeado por una "roseta" de neutrófilos (*abajo*).

minantes antigénicos idénticos a los de proteínas de la glándula tiroidea del conejo. El mecanismo de la enfermedad autoinmunitaria consistiría en un ataque directo producido por el aparato inmunitario del organismo, probablemente a cargo de anticuerpos y células *T* contra los propios tejidos.

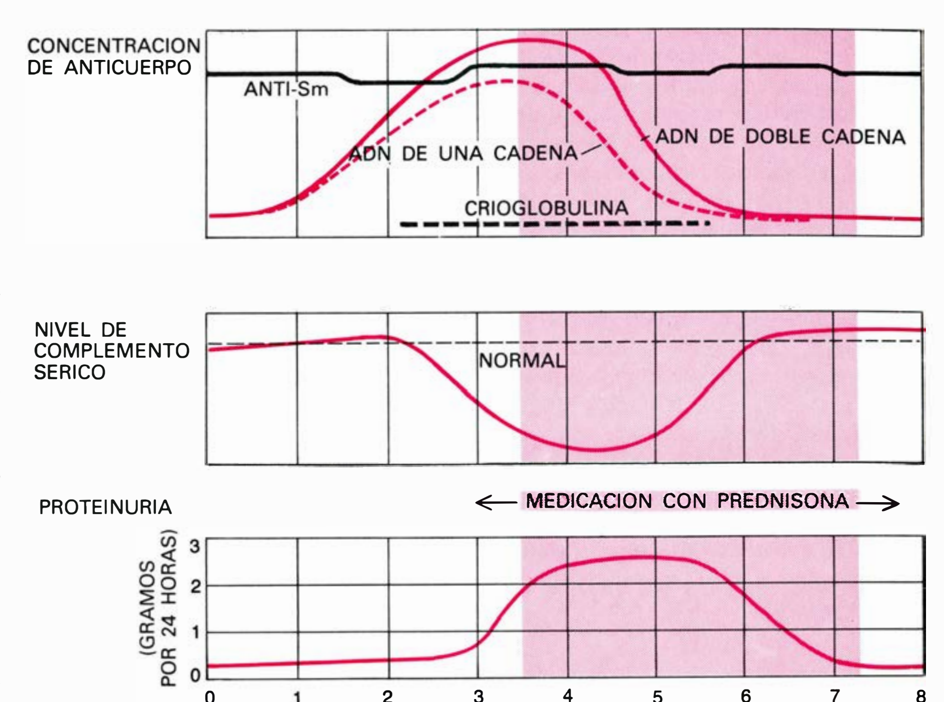
Existe otro mecanismo de autolesión a través del sistema inmunitario. Se descubrió en la primera década del presente siglo, cuando Clemens von Pirquet, de la Facultad de Medicina Johns Hopkins, especulaba sobre la causa de la "enfermedad del suero" observada en personas que habían sido tratadas con un antisuero preparado en caballos para contrarrestar, por ejemplo, los efectos de la toxina tetánica. Von Pirquet sugirió que la enfermedad del suero se producía como consecuencia de la interacción de los anticuerpos del paciente con antígenos extraños (proteínas del suero del caballo), que provocaban la formación de una sustancia tóxica. Estaba en lo cierto; la sustancia era un complejo inmunitario y era responsable de la erupción, dolores articulares y otros síntomas de la enfermedad del suero. El mecanismo de la lesión tisular en estas enfermedades supone la formación de complejos antígeno-anticuerpo tóxicos, en el curso de la respuesta inmunitaria a un antígeno.



RESPUESTA INMUNITARIA llevada a cabo por anticuerpos circulantes y células sensibilizadas. Las células precursoras de la médula ósea (*1*) se diferencian dando lugar a células linfoides, que migran a tejidos linfoides asociados con el intestino (*2a*) y el timo (*2b*), donde maduran para dar lugar a los linfocitos *B* y *T* (*3*), respectivamente. Una célula *B* está programada genéticamente para producir unas cuantas copias de un determinado anticuerpo, que actúa como receptor en la superficie celular. Los diversos tipos de células *T*, con receptores similares, tienen funciones reguladoras o resultan sensibilizadas por los antígenos para actuar en la inmunidad celular. Existen cinco clases de antígenos, de los cuales las IgM y las IgG son los de mayor importancia en el LES. En él, una célula productora de IgM pasa a sintetizar immunoglobulina gamma (IgG) (*4*). Un determinante antigénico (un punto concreto de la molécula del antígeno) interacciona con el sitio de unión de una IgG receptora, lo que induce la formación de un clon de células plasmáticas (*5*) a partir de las células *B*. Estas células plasmáticas sintetizan grandes cantidades de un anticuerpo específico, que puede identificarse por su idiotipo característico, un determinante antigénico localizado en el sitio de combinación del anticuerpo. El mismo idiotipo puede estar presente en los receptores de la superficie de las células *T* que modulan la respuesta de los anticuerpos. El anticuerpo puede reaccionar con un sitio del mismo antígeno que estimuló originalmente la proliferación de las células *B;* la unión del antígeno con el anticuerpo forma el complejo (*6*).

Tanto los autoanticuerpos como los complejos inmunitarios son ubicuos en el suero de los mamíferos. Su presencia no es la consecuencia de un accidente, sino que aparecen como componentes fisiológicos del sistema inmunitario y desarrollan un importante papel en el mantenimiento de la homeostasis, la estabilidad del medio interno. De acuerdo con la teoría propuesta por Niels Kaj Jerne, del Instituto de Inmunología de Basilea, un determinado tipo de autoanticuerpos, los anticuerpos anti-idiotipo, servirían para autorregular el sistema inmunitario [*véase* "The Inmune System", por Niels Kaj Jerne; SCIENTIFIC AMERICAN, julio de 1973]. Se trata de anticuerpos que actúan frente a sitios antigénicos (llamados idiotipos) que se encuentran en el lugar de combinación de la molécula IgG o en sus proximidades y que son específicos para anticuerpos producidos por un clon único de células plasmáticas. Tales sitios antigénicos fueron descritos por primera vez en el hombre por Henry G. Kunkel, de la Universidad Rockefeller, y en animales por Jacques Oudin, del Instituto Pasteur. En animales de experimentación se han identificado anticuerpos anti-idiotipo. La inducción de tales anticuerpos incrementa o bloquea la respuesta inmunitaria del animal.

Entre los autoanticuerpos más comunes en el hombre se cuentan los "factores reumatoides", grupo de anticuerpos dirigidos frente a las globulinas gamma. Los factores reumatoides están presentes en pequeñas concentraciones en personas normales y en concentraciones mucho mayores en pacientes de diversas enfermedades, particularmente en la artritis reumatoide y en el LES, donde pueden desempeñar un importante papel en el desarrollo de la enfermedad. Otros autoanticuerpos están dirigidos contra los componentes del núcleo celular. Se encuentran en pequeña proporción en individuos normales, sobre todo en personas de edad avanzada y, como se indicará más adelante, en pacientes con LES. (Debe señalarse que la distinción entre niveles



EXACERBACIONES DE LA ENFERMEDAD en un paciente con LES y manifestaciones renales. Los niveles séricos de anticuerpos frente al ADN de doble cadena y de cadena única aumentan, y el nivel del complemento empieza a declinar, antes del comienzo de la proteinuria (incremento de la eliminación urinaria de proteínas), que constituye un signo de alteración renal. Obsérvese la rápida disminución de los anticuerpos asociada a una caída extrema del complemento. Los anticuerpos se han combinado con los antígenos formando complejos inmunitarios que fijan el complemento (los complejos pueden observarse en el suero en forma de crioglobulinas, un precipitado geliforme que aparece a bajas temperaturas). Otro anticuerpo antinuclear, el anti-Sm, no aumenta en las exacerbaciones de la enfermedad. Aparentemente, sólo determinados anticuerpos forman complejos inmunitarios capaces de inducir lesiones tisulares. La prednisona, un esteroide antiinflamatorio, se muestra eficaz en el tratamiento de los episodios agudos. La enfermedad renal puede limitarse a una glomerulonefritis focal, pero a veces desemboca en una fase difusa.

"normales" y "anormales" depende de la sensibilidad del método empleado en la determinación de los anticuerpos y resulta cada vez más difícil a medida que se dispone de pruebas inmunológicas de mayor precisión.)

¿Cuáles son las funciones biológicas de los autoanticuerpos y de los complejos inmunitarios normalmente presentes en ausencia de una agresión por agentes extraños? Todas las pruebas de que disponemos en la actualidad sugieren que desempeñan un papel central en el mantenimiento de la homeostasis. Los fenómenos asociados al LES representan alteraciones funcionales de los mecanismos homeostáticos normales. Tales alteraciones se expresan clínicamente por un grupo de síntomas que constituyen el síndrome LES.

El lupus eritematoso fue identificado a mediados del siglo XIX como una enfermedad limitada a la piel, pero a finales del mismo, Moritz Kaposi, en Viena, y William Osler, en la Universidad de Johns Hopkins, describieron una serie de manifestaciones sistémicas que aparecían a menudo asociadas al "lupus facial": artritis, neumonía y síntomas del sistema nervioso central. A principios de nuestro siglo se describió la patología de varios órganos abdominales como parte del cuadro clínico. En 1941, el LES fue clasificado por Paul Klemperer, Abou D. Pollack y George Baehr, del Hospital Monte Sinaí de Nueva York, como una enfermedad "colágeno-vascular", una enfermedad que afectaba al tejido conectivo y a los vasos sanguíneos.

Fue el descubrimiento, en 1948, de la célula del lupus eritematoso (LE) por M. M. Hargraves, de la Clínica Mayo, lo que dirigió la atención hacia los aspectos inmunológicos del LES. Halsted R. Holman y Kunkel demostraron que el suero de pacientes de LES contenía una globulina gamma con afinidad para los núcleos, y específicamente para las nucleoproteínas, proteínas unidas al ADN en los cromosomas. En el suero de los pacientes hallaron varios anticuerpos capaces de reaccionar con el núcleo, y observaron que uno de ellos, un anticuerpo frente al complejo ADN-proteína, era responsable del fenómeno de la célula LE. La observación de que la mayoría de los sueros de pacientes con LES contenían anticuerpos frente a los propios antígenos del paciente, hizo pensar que se trataba de una enfermedad autoinmunitaria.

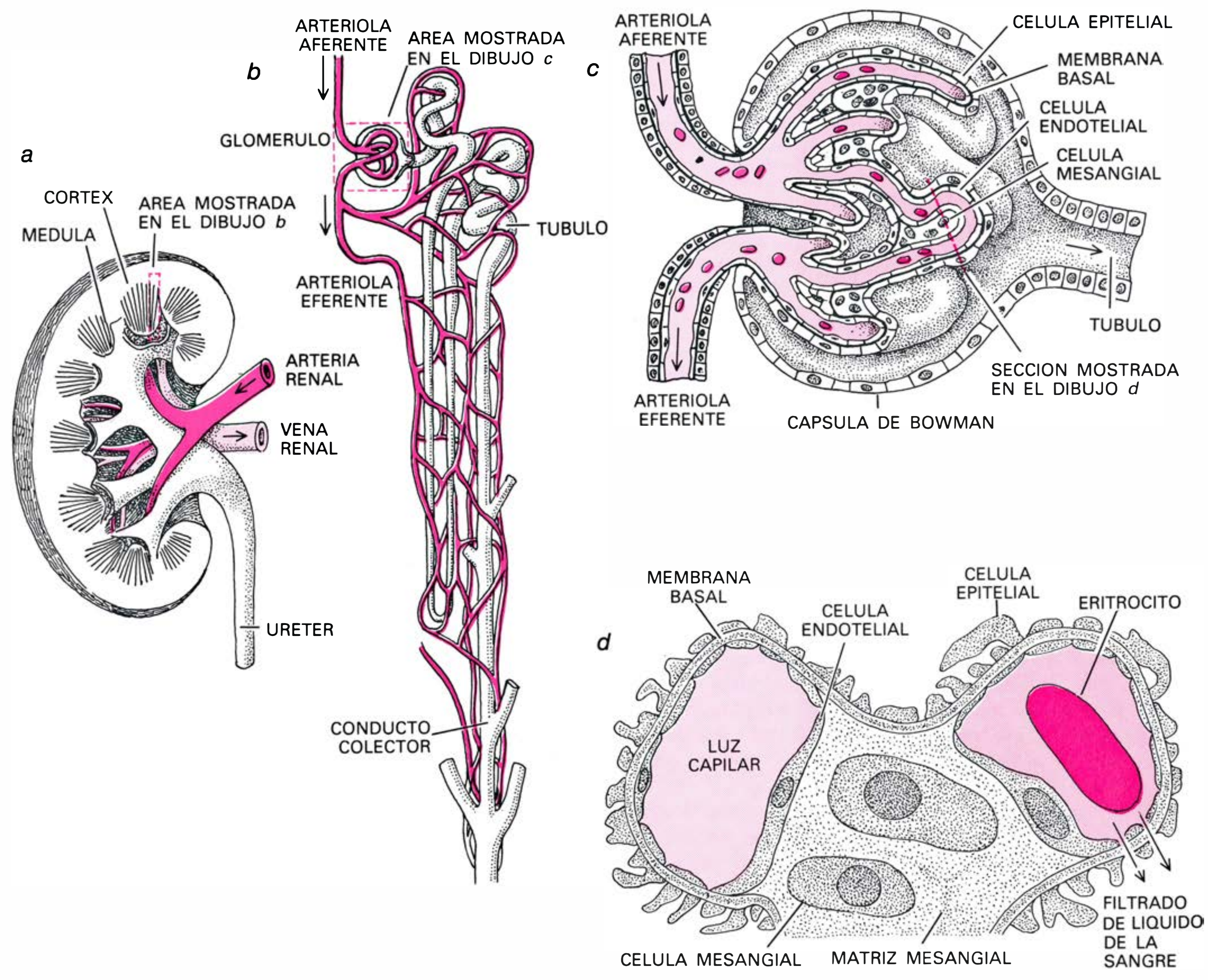
Es difícil, no obstante, invocar las reacciones autoinmunitarias para explicar todos los síntomas de los pacientes con LES: fiebre, fatiga, erupción cutánea, artritis, inflamación vascular, pleuresía, pericarditis y glomerulonefritis. Esta última, una enfermedad del riñón, se caracteriza por la inflamación de los glomérulos renales (que filtran la sangre) y es una de las manifestaciones más graves del LES. Los anticuerpos antinucleares, un tipo muy común de autoanticuerpos, no atraviesan normalmente la membrana de las células vivas intactas. Habitualmente no se detectan depósitos de anticuerpos antinucleares en los tejidos de los pacientes con LES, como sería de esperar si tales anticuerpos fueran responsables de la patología.

Las manifestaciones clínicas y la anatomía patológica del LES se parecen mucho más a la enfermedad del suero que a las enfermedades autoinmunitarias típicas. Recordemos que la enfermedad del suero resultaba de la formación de complejos inmunitarios. Diferentes tipos de experimentos sugieren un papel patogénico directo de los complejos antígeno-anticuerpo. Al mismo tiempo que se descubrían los anticuerpos antinucleares circulantes en el suero de pacientes con LES, Robert C. Mellors, de la Facultad de Medicina de la Universidad Cornell, observó depósitos de gamma globulina en los glomérulos renales de dichos pacientes y Jacinto J. Vázquez y Frank J. Dixon, entonces en la Facultad de Medicina de la Universidad de Pittsburgh, en las pequeñas arterias del hígado y del bazo, así como en las células LE. Los anticuerpos circulantes se dirigían principalmente contra antígenos nucleares: ADN, tanto en su forma nativa de doble cadena como en la forma desnaturalizada de cadena única, y contra los complejos de ADN e histonas. Eng M. Tan, Kunkel y colaboradores mostraron que los niveles séricos de los antígenos de ADN y de los correspondientes anticuerpos fluctuaban en distintos pacientes, sugiriendo una relación entre la formación de complejos inmunitarios y el incremento de los síntomas de la enfermedad. Estudios ulteriores han demostrado una sorprendente correlación entre las exacerbaciones de la enfermedad y la reducción del nivel sérico del complemento, que es fijado por los complejos antígeno-anticuerpo y depositado en los tejidos o eliminado por las células reticuloendoteliales.

Trabajando con Kunkel y otros en su laboratorio, logré una prueba visual directa de que los complejos antígeno-anticuerpo podían ser importantes mediadores tanto de la respuesta vascular como de las lesiones renales en el LES, dependiendo del sitio de deposición. Examinamos biopsias de riñón en distintos estadios de la enfermedad. Utilizando sistemas de tinción convencional, el microscopio óptico mostraba diferentes grados de lesión tisular. Con el microscopio de fluorescencia podían observarse, en preparaciones tratadas con un anticuerpo marcado con un colorante fluorescente, depósitos de complejos inmunitarios en el mesangio, región del glomérulo que elimina macromoléculas de la sangre. Estos depósitos aparecerían en la mayoría de los pacientes con LES.

Si los depósitos se limitaban al mesangio, la proteinuria (exceso de proteína en la orina, el signo principal de glomerulonefritis) era mínima o inexistente. Las biopsias de pacientes con proteinuria importante mostraban depósitos granulares de gamma globulina junto al complemento a lo largo de la membrana basal, la membrana a través de la cual se filtra la sangre. El depósito continuado de dichos complejos determina el engrosamiento de las paredes capilares. La observación de tal engrosamiento en los glomérulos renales de pacientes con LES condujo a Klemperer, en 1935, a relacionar el LES con las lesiones de la enfermedad del suero que se apreciaban en los riñones de caballos que habían recibido inyecciones repetidas de toxoide diftérico para preparar antisuero.

El papel de los complejos en el LES permaneció sin aclarar hasta que en la década de los 50, Frederick G. Germuth, Jr., entonces en la Universidad de Johns Hopkins, realizó una serie de experimentos con animales, que demostraron su importancia y que relacionaron específicamente la gravedad de la enfermedad con la respuesta inmunitaria a un antígeno. Germuth y colaboradores mostraron que la administración, a largo plazo, de antígenos proteicos a conejos desencadenaba una enfermedad crónica similar a la enfermedad del suero en animales y provocaba lesiones glomerulares parecidas a las observadas en los riñones de pacientes con LES. Cuando los conejos se inmunizaban con una dosis constante de una proteína (albúmina de suero bovina), la gravedad de la glomerulo-



GLOMERULONEFRITIS, inflamación de las asas capilares de los glomérulos renales. Esta alteración constituye una manifestación grave del LES. El riñón (*a*) es el órgano a través del cual se eliminan las sustancias y fluidos tóxicos de la sangre, dando lugar a la formación de la orina. Este órgano está constituido por alrededor de un millón de unidades funcionales denominadas nefronas. Cada nefrona (*b*) consta de un manojo de asas capilares, el glomérulo, donde se filtra la sangre y un túbulo a través del cual el filtrado pasa al tubo colector. Gran parte del flúido y algunas sustancias disueltas se reabsorben pasando a la red capilar que rodea al túbulo. El glomérulo (*c*) está constituido por una serie de capilares interconectados. Una sección del glomérulo (*d*) muestra la luz capilar, rodeada por una membrana basal cubierta por células epiteliales. La membrana y las células epiteliales controlan la filtración. Las macromoléculas se eliminan de la sangre y se depositan en el mesangio. El depósito de complejo altera la función renal.

nefritis dependía de la respuesta de anticuerpos característica de cada conejo. En aquellos conejos en los que se producía una respuesta potente (grandes cantidades de anticuerpo), se formaban extensas tramas de complejo antígeno-anticuerpo que eran fácilmente eliminadas por el sistema reticuloendotelial, y que aparentemente no eran tóxicas. Los conejos con una respuesta de anticuerpos inferior formaban complejos más pequeños, con afinidad por el mesangio. Los conejos con una respuesta aún más débil producían complejos todavía menores y desarrollaban una glomerulonefritis proliferativa difusa, la forma grave de la enfermedad. Respuestas de menor intensidad en la producción de anticuerpos daban lugar a complejos menores incluso, y no se acompañaban de signos patológicos.

Estudios previos llevados a cabo por Theodore P. Pincus y Charles L. Christian, de la Facultad de Medicina y Cirugía de la Universidad de Columbia, indicaban que las propiedades intrínsecas de los anticuerpos frente a la albúmina de suero bovina influían en la toxicidad de los complejos. Germuth, partiendo de investigaciones realizadas en la Facultad de Medicina de la Universidad de San Luis, hizo énfasis en diferencias cualitativas del anticuerpo, tales como su avidez o tendencia a unirse al antígeno. Dixon y colaboradores, que trabajan actualmente en la Clínica Scripps, realizaron una serie de importantes experimentos en los que se ajustaba la dosis de antígeno de forma que se mantuviera un moderado exceso de antígeno, es decir, un estado en el que había algo más de antígeno que de anticuerpo en los complejos. Estos complejos eran especialmente aptos para provocar una glomerulonefritis membranosa crónica. El comienzo de la nefritis coincidía con un marcado incremento de depósitos en el riñón, lo que sugería que la lesión inicial facilitaba el incremento de los depósitos y que un posible factor coadyuvante era la saturación del sistema reticuloendotelial (que no era entonces capaz de eliminar los complejos). Robert T. McCluskey y Baruj Benacerraf, a la sazón en la Facultad de Medicina de la Universidad de Nueva York, lograron producir glomerulonefritis y vasculitis (inflamación de los vasos sanguíneos) en ratones por administración directa de complejos inmunitarios solubles, preparados en el laboratorio.

En definitiva, los complejos inmunitarios varían en cuanto a su potencial tóxico y ello explica en parte el amplio margen de síntomas clínicos. El lugar del depósito y la iniciación de la respuesta inflamatoria dependen del tamaño del complejo, del número de sitios de unión de los antígenos con los anticuerpos, del número de anticuerpos distintos y de su afinidad por los antígenos. Charles G. Cochrane, de la Clínica Scripps, observó además que los agentes que tendían a incrementar la permeabilidad de los vasos sanguíneos afectaban también a la toxicidad de los complejos. El tratamiento de conejos con antagonistas de agentes tales como la histamina y serotonina reducía el depósito de complejos y la consiguiente inflamación. Mart Mannik, de la Facultad de Medicina de la Universidad de Washington, demostró que la capacidad del sistema reticuloendotelial para eliminar complejos se veía afectada por la capacidad de los fagocitos para reconocer una región determinada del anticuerpo.

En algunos pacientes de LES con síntomas renales, la enfermedad se limita a una glomerulonefritis focal en la que sólo algunos glomérulos tienen complejos inmunitarios y sufren lesiones. Drogas que reducen la inflamación o suprimen la respuesta inmunitaria parecen ser particularmente eficaces en este estadio de la enfermedad. En otros pacientes, la enfermedad progresa hacia una fase difusa proliferativa con depósitos masivos en la mayoría de glomérulos e importante proteinuria. Recientemente, Giuseppe A. Andres, de la Facultad de Medicina de la Universidad del estado de Nueva York en Buffalo, y McCluskey, que se encuentra actualmente en la Facultad de Medicina de Harvard, han hecho énfasis en la importancia de las lesiones del túbulo renal (donde el agua y las sustancias disueltas se reabsorben a partir del filtrado glomerular) como consecuencia del depósito de complejo en el tejido que rodea el túbulo.

En un esfuerzo dirigido a la caracterización de la naturaleza de los complejos detectados en formas progresivas de la enfermedad, Kunkel, Peter H. Schur y el autor examinaron riñones de pacientes que habían fallecido de la enfermedad renal del LES; en colaboración con John B. Winfield, el autor examinó las crioglobulinas (complejos inmunitarios precipitados a baja temperatura a partir del suero) de pacientes con LES. De los glomérulos se eluyeron (lavaron) varios anticuerpos y se aislaron anticuerpos similares a partir de las crioglobulinas. En ambas localizaciones, los anticuerpos se encontraban a mayor concentración que en el suero, sugiriendo que se habían concentrado mediante la formación de complejos antígeno-anticuerpo. Los anticuerpos se ensayaron frente a ADN de doble cadena, ADN de cadena única y ribonucleoproteínas; los complejos que contenían ADN eran los más importantes, y el antígeno ADN pudo detectarse en glomérulos que contenían anticuerpos anti-ADN.

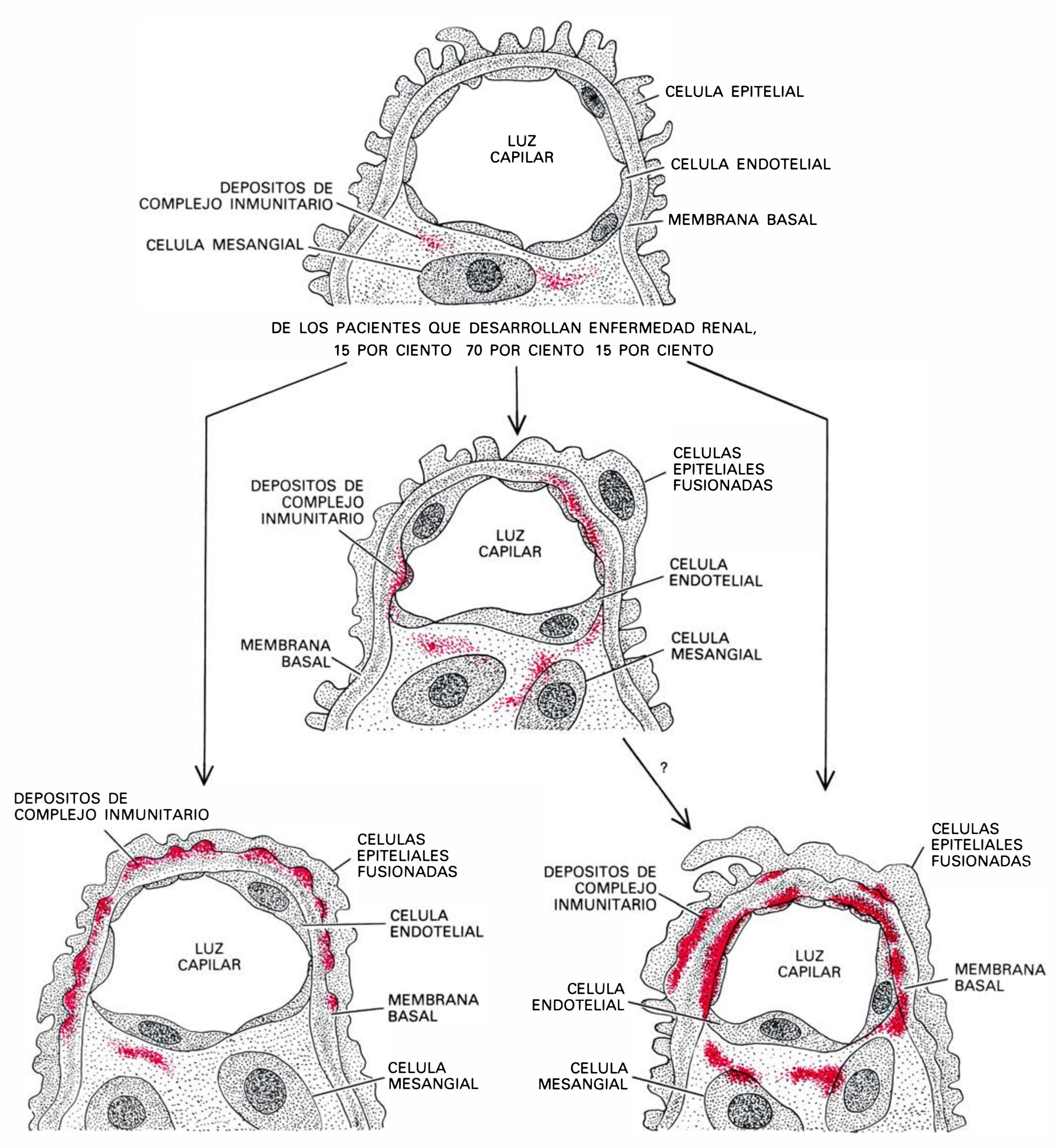
El complejo anticuerpo-ADN fue el primer agente específico productor de lesión tisular identificado en la glomerulonefritis humana; otros antígenos se han localizado en los riñones de pacientes que sufrían glomerulonefritis "primaria", una enfermedad renal similar a la del LES, pero no asociada con él. El LES ha servido de prototipo para el estudio de las enfermedades producidas por complejos inmunitarios.

Winfield, Kunkel y el autor observaron que los anticuerpos anti-ADN eluidos de glomérulos tenían una mayor afinidad que los mismos anticuerpos obtenidos del suero del paciente. Es más, los anticuerpos circulantes anti-ADN de pacientes con enfermedad renal tenían menos afinidad que los propios de pacientes sin enfermedad renal. Estos hechos sugieren que, en la glomerulonefritis, los anticuerpos de alta afinidad se depositan en el riñon preferencialmente en forma de complejos. En consecuencia, si se determinan únicamente los anticuerpos circulantes presentes en el suero, no se detectarán aquellos anticuerpos que formen parte de un complejo. La posibilidad de identificar complejos circulantes y de evaluar la actividad del proceso patológico se ha visto facilitada por varios descubrimientos; entre ellos, el aislamiento de los componentes del complemento y la caracterización de las vías del complemento llevados a cabo por Hans J. Müller-Eberhard, de la Clínica Scripps, por Shaun Ruddy, de la Facultad de Medicina de Virginia, y por otros investigadores.

Una forma de abordar el estudio de los complejos consiste en examinar las crioglobulinas, cuyas interacciones con el complemento han sido estudiadas por Christian. Vincent Agnello, Kunkel y el autor observaron que pacientes con grandes cantidades de crioglobulinas mostraban importantes depósitos de IgM en sus riñones, en contraste con la mayoría de pacientes con LES renal en los que la principal inmunoglobulina era la IgG. El factor reumatoide IgM es también el principal componente de las

crioglobulinas circulantes en el LES. Estos hechos indican que los complejos de globulinas gamma y anticuerpos frente a ellas contribuyen a las lesiones renales en al menos un grupo de pacientes. Recientemente, Morris Reichlin, de la Facultad de Medicina de la Universidad estatal de Nueva York en Buffalo, ha observado otro tipo de complejo inmunitario que implica antígenos del citoplasma de las células, lo que sugiere, de nuevo, que varios tipos de complejos pueden contribuir a la lesión glomerular.

La observación realizada por Agnello, Robert J. Winchester y Kunkel de que la C1q (proteína del complemento que interacciona con el anticuerpo) forma un precipitado con agregados de inmunoglobulina, proporciona un sistema para detectar complejos circulantes. Agnello, Kunkel y el autor han observado tales precipitados en el suero de pacientes con LES. Paul H.



DEPOSITO DE COMPLEJO INMUNITARIO en el riñón, esquematizado en diferentes fases del LES. Incluso un paciente sin afectación renal, con glomérulos morfológicamente normales, muestra depósitos de complejo inmunitario en el mesangio (*arriba*). Dos tercios de los pacientes muestran manifestaciones renales y un 70 por ciento de los mismos desarrolla glomerulonefritis proliferativa focal y proteinuria ligera o moderada (*dibujo central*), con depósitos dispersos a lo largo de la membrana basal, debajo de la capa endotelial. Un 15 por ciento desarrollan glomerulonefritis membranosa con proteinuria moderada o intensa (*abajo, a la izquierda*). El depósito continuado de complejo determina su distribución uniforme a lo largo de la membrana basal. Otro 15 por ciento (y también una cierta proporción de los que desarrollan enfermedad focal) sufren una glomerulonefritis proliferativa difusa, con depósitos masivos de complejos inmunitarios en la mayoría de glomérulos (*abajo, a la derecha*), que suele conducir a una insuficiencia renal.

Lambert y Peter A. Miescher, de la Universidad de Ginebra, han desarrollado otro ensayo para complejos en el que se utiliza la C1q marcada con un isótopo radiactivo. Y Argyrios N. Theofilophoulos y Dixon han ideado un ensayo con células linfoides (denominadas células Rajii) que poseen receptores para componentes del complemento unidos a complejos. Estas y otro tipo de pruebas que detectan complejos inmunitarios proporcionan información sobre el curso del LES que no puede obtenerse por la mera determinación del nivel de antígeno o anticuerpo.

Aunque se ha establecido una clara relación entre los complejos inmunitarios y las lesiones de la enfermedad renal y vascular, resulta difícil identificar el mecanismo de las alteraciones tisulares en otras manifestaciones del LES. Los estudios realizados en biopsias cutáneas por inmunofluorescencia, por citar un ejemplo, muestran que alrededor de la mitad de los pacientes de LES poseen depósitos de complejos inmunitarios en la piel justo por debajo de la capa más externa (la epidermis), pero los depósitos no producen siempre inflamación. Tan y Kunkel observaron depósitos granulares de gamma globulina tanto en lesiones inflamatorias como en áreas cutáneas normales. Morris Ziff y James N. Gilliam, del Centro de Ciencias de la Salud de la Universidad de Texas en Dallas, y Reichlin hallaron que la incidencia de tales depósitos en la piel normal aumentaba en pacientes con enfermedad renal, los cuales, a menudo, mostraban mayores cantidades de complejos inmunitarios circulantes. Los depósitos pueden hallarse también en el LES "discoide", en el que los cambios cutáneos son el síntoma principal. Existen múltiples cuestiones sin resolver sobre tales lesiones y los depósitos asociados. ¿Qué diferencia el efecto de los complejos en la piel normal y en la piel inflamada: las propiedades de los depósitos en sí o la reacción del paciente frente a los mismos? ¿Son distintos los complejos depositados en la piel respecto de los depositados en los glomérulos? Y si son diferentes, ¿por qué?

Se sabe muy poco del mecanismo de los síntomas neurológicos del LES, que se manifiestan únicamente al cabo de varios años del comienzo de la enfermedad y que son cada vez más frecuentes a medida que se alarga la vida media de estos pacientes. Hay escasos datos que establezcan relación entre los depósitos y lesiones cerebrales, e incluso entre determinadas lesiones y síntomas tales como las convulsiones y las psicosis. Se han detectado varios anticuerpos en el suero de pacientes con LES que reaccionan con antígenos del sistema nervioso, pero no se han relacionado con la enfermedad neurológica. Las pruebas aducidas sobre un efecto directo de los autoanticuerpos (a diferencia de los complejos) en el LES implican fundamentalmente la interacción de los anticuerpos circulantes con los elementos celulares de la sangre. La reacción de anticuerpos con antígenos de la superficie de los glóbulos rojos puede provocar anemia hemolítica; los eritrocitos están alterados y el nivel de hemoglobina se reduce.

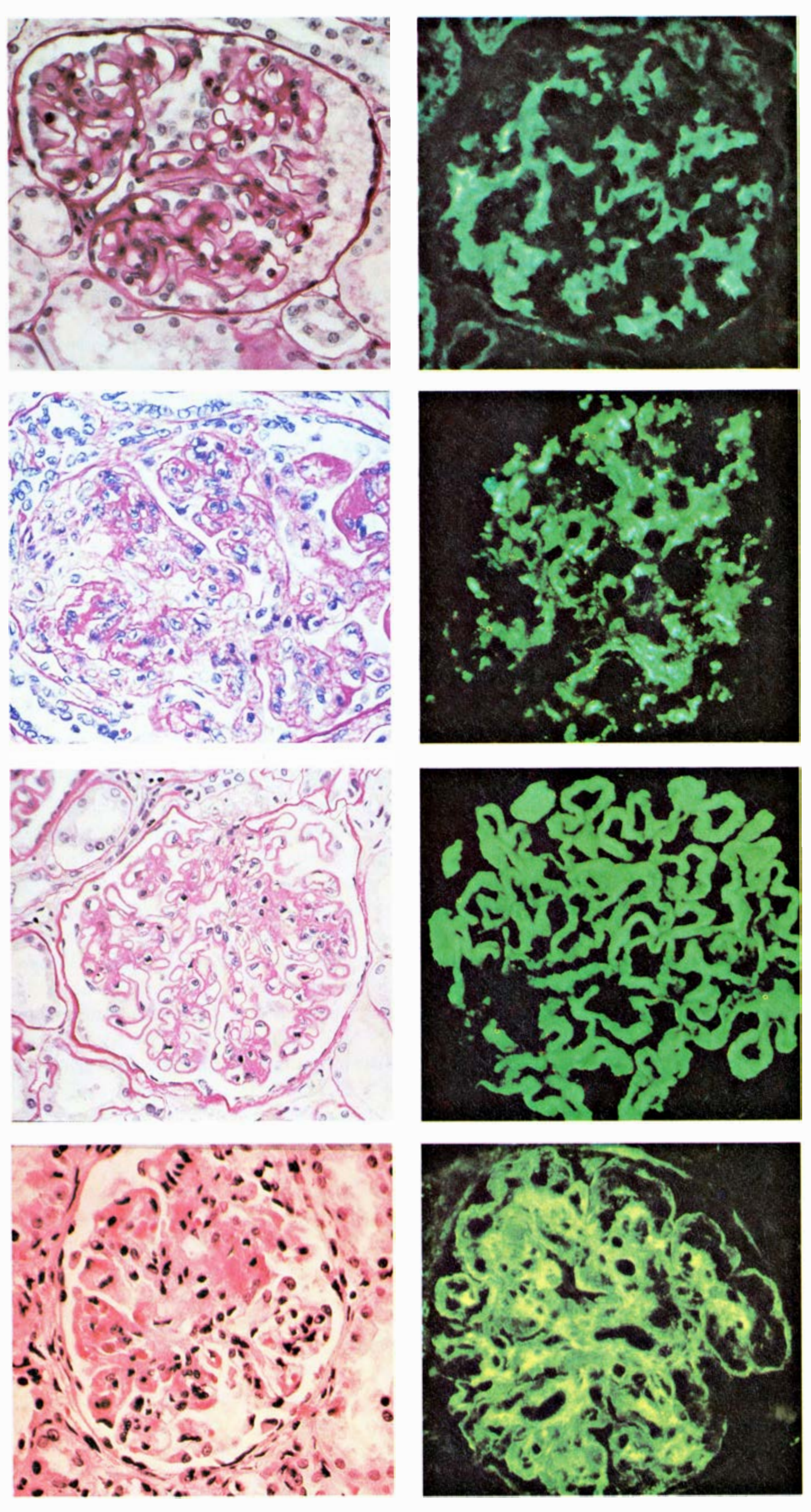
En el suero de pacientes con LES se detectan anticuerpos frente a linfocitos *B* y *T*. Se han observado alteraciones en la función de las células *T*; entre ellas una reducción de la actividad celular supresora. Una población de anticuerpos frente a células *T* muestra reacción cruzada con antígenos similares a antígenos del cerebro del ratón, pero no se sabe si tales anticuerpos alteran la función inmunitaria celular o si afectan al tejido cerebral. Winfield, Winchester y Kunkel observaron que la presencia de ciertos anticuerpos antilinfocitos en el suero de pacientes con LES estaba asociada a un bajo nivel de linfocitos circulantes. También detectaron altas concentraciones de tales anticuerpos en las crioglobulinas, lo que sugiere que deben formar complejos inmunitarios con antígenos superficiales liberados por los linfocitos circulantes.

Con excepción de los anticuerpos responsables de la anemia hemolítica, no se ha demostrado claramente la existencia de autoanticuerpos productores de las lesiones tisulares del LES. Actualmente se acepta que los complejos inmunitarios causan las lesiones en los vasos sanguíneos y en los glomérulos. Los estudios realizados por Andres en la membrana sinovial de las articulaciones, en la pleura y en el pericardio (las membranas que rodean al pulmón y al corazón, respectivamente) y en los septos alveolares (las paredes de los alveolos pulmonares), sugieren que los complejos producen muchos de los síntomas característicos del LES.

Hasta ahora se han discutido los posibles mecanismos implicados en las lesiones tisulares del LES. Decir que esta enfermedad está ligada a la formación de complejos inmunitarios no significa identificar la causa o causas que la producen. Esencialmente desconocemos la etiología, habiéndose establecido únicamente algunos hechos. Existe una predisposición genética para sufrir la enfermedad, y en el proceso están implicadas anomalías del sistema inmunitario. Por otra parte, otros factores podrían estar relacionados con la enfermedad, entre ellos ciertas drogas, estrógenos, radiaciones ultravioleta e infecciones víricas.

El hecho más notable en favor de la importancia del factor genético en la enfermedad es la observación de que si uno de dos gemelos univitelinos sufre la enfermedad, el otro la padece también en un 50 a 60 por ciento de los casos, mientras que la concordancia es del 2 al 3 por ciento en gemelos bivitelinos y en familiares de primer grado. La incidencia en el conjunto de la población es sólo del 0,2 por ciento. Varias alteraciones congénitas, raras, de las proteínas del complemento parecen relacionadas con el LES. Genes que controlan la síntesis de ciertas proteínas del complemento están en el mismo cromosoma que otros que controlan la respuesta inmunitaria y la producción de antígenos de histocompatibilidad, un conjunto de moléculas que difieren de un individuo a otro y se reconocen en el rechazo de injertos. La incidencia del LES está marcadamente aumentada en personas con ciertos tipos de histocompatibilidad. Si bien todas estas observaciones indican la existencia de una influencia genética en el LES, se desconoce por qué mecanismos actúan los factores genéticos en la enfermedad. La relación entre el factor genético y el LES podría depender de algún factor coadyuvante. Así, un defecto en el sistema de complemento o en el sistema inmunitario podrían significar una predisposición hacia una infección vírica que a su vez sería la causa de la enfermedad.

En el LES se han detectado alteraciones funcionales de los anticuerpos y de los componentes celulares del sistema inmunitario. Los pacientes poseen diferentes tipos de autoanticuerpos con una tendencia particular hacia la formación de complejos inmunitarios. Uno de tales anticuerpos, el anticuerpo frente al ADN de doble cadena, se encuentra prácticamente de forma exclusiva en el LES. Existe un subgrupo de anticuerpos anti-ADN capaces de formar complejos que fijan el complemento y determinar la destrucción tisular. El hecho poco frecuente de la presencia de tales anticuerpos en otras enfermedades, y la imposibilidad de inducir anticuerpos inmunizando ani-



PREPARACIONES DE TEJIDO que muestran cuatro fases de la enfermedad renal con los cambios estructurales puestos de manifiesto por la tinción histológica (*izquierda*) y la localización de los complejos inmunitarios observada con el microscopio de fluorescencia (*derecha*). El glomérulo estructural y funcionalmente normal observado en una biopsia (*arriba, izquierda*) muestra depósitos irregulares filiformes de complejos inmunitarios en áreas mesangiales (*arriba, derecha*). Un glomérulo de un paciente con enfermedad renal focal (*el segundo empezando por arriba*) muestra un engrosamiento irregular de la membrana y cierta fusión de los manojos capilares; depósitos granulares irregulares de complejos están unidos a las paredes capilares. En la enfermedad membranosa (*tercera, empezando por arriba*), la membrana parece más engrosada (*izquierda*), lesión que se denomina "fenómeno del asa de alambre" (*derecha*). En la enfermedad difusa (*abajo*) se observan alteraciones estructurales (*izquierda*) y depósitos masivos de complejo inmunitario (*derecha*) que pueden obliterar los glomérulos alterando la filtración.

males con ADN de doble cadena, sugiere que el sistema inmunitario no reconoce en circunstancias normales al ADN como sustancia extraña. No obstante, Arthur D. Bankhurst y Ralph C. Williams, Jr., de la Facultad de Medicina de la Universidad de Nuevo México, observaron en individuos normales linfocitos *B* circulantes con receptores para el antígeno ADN, siendo su concentración superior en pacientes con LES. Da la impresión de que las líneas celulares capaces de producir anticuerpos frente a este antígeno particular han logrado escapar del control del sistema inmunitario en los pacientes con LES.

¿Cómo puede suceder ese fenómeno? Se han sugerido varias explicaciones a partir de los datos obtenidos en estudios sobre mecanismos de inmunidad celular. Tanto los pacientes con LES, como una cepa de ratones con una enfermedad parecida en ciertos aspectos al LES, muestran una deficiencia en la función supresora de las células *T*. La función supresora disminuye con la edad en ratones que padecen un síndrome parecido al LES. La disminución está asociada a un incremento de la producción de anticuerpos anti-ADN y a una forma virulenta de glomerulonefritis. Alfred D. Steinberg, Anthony S. Fauci, M. Eric Gershwin y Thomas A. Waldmann, de los Institutos Nacionales de la Salud, han obtenido datos que ponen de manifiesto la existencia de defectos en la función supresora celular, así como sobre el funcionamiento de las células *B* en ratones y en pacientes humanos con LES.

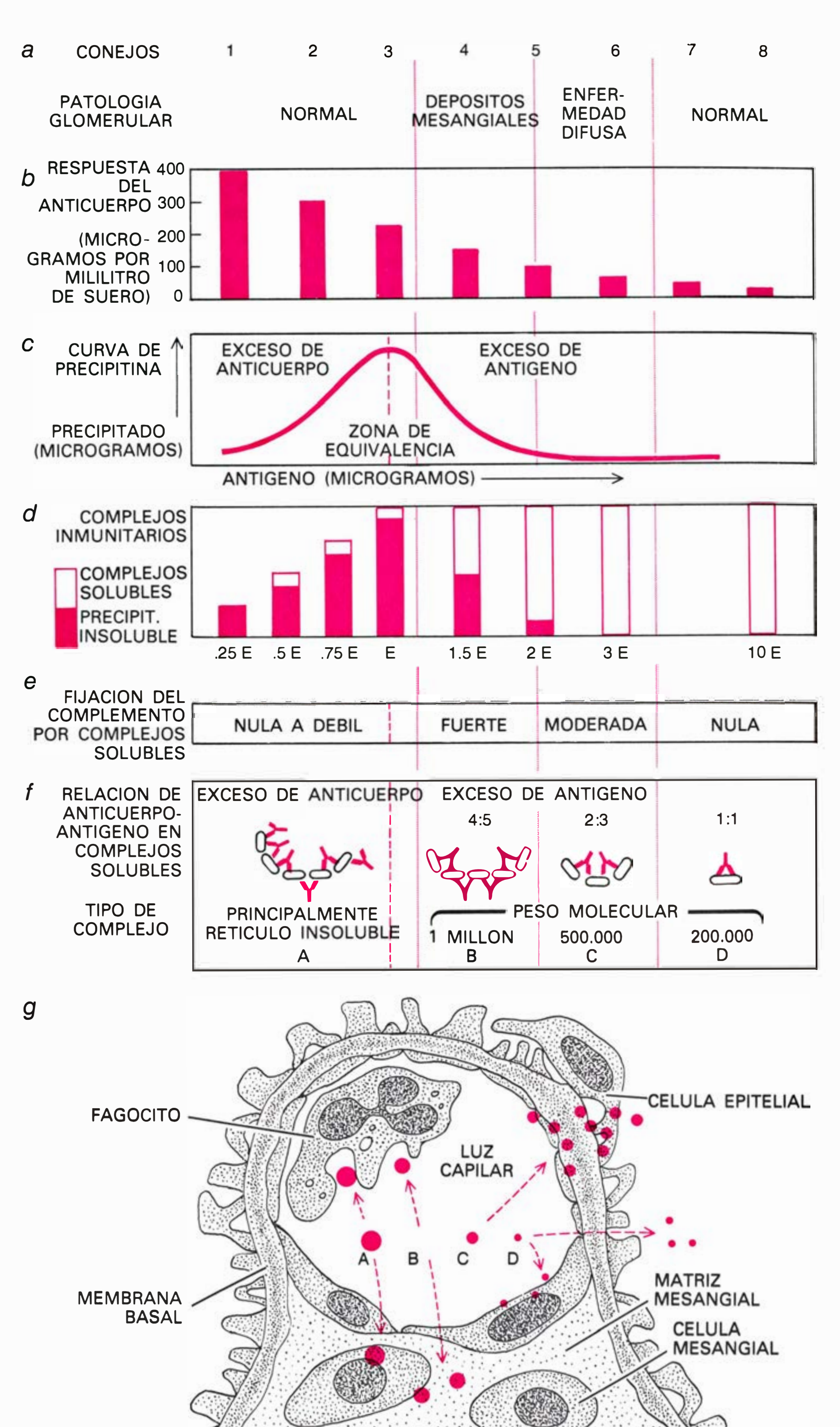
En el hombre, el nivel de anticuerpos antinucleares aumenta con la edad, quizá como consecuencia de una reducción en la función celular supresora, no habiéndose atribuido síntomas a su incremento. Se ha sugerido que ciertos anticuerpos frente a células *T* que se detectan en el hombre y en el ratón contribuyen a las anomalías observadas en las células supresoras. Peter Wernet y Kunkel demostraron la presencia de anticuerpos que interaccionaban con las células *T* e inhibían su proliferación. Tal inhibición, si afectara a células supresoras del organismo, liberaría a las células *B* productoras de anticuerpos de la acción supresora. Robert S. Schwartz, de la Facultad de Medicina de la Universidad de Tufts, ha detectado recientemente anomalías en las células supresoras a familiares de pacientes con LES. Por otra parte, es posible que algún factor, tal como la interacción con complejos inmunitarios circu-

lantes, determine la proliferación de las células coadyuvantes y, en consecuencia, aumente la producción de anticuerpos por los linfocitos *B*.

Los datos obtenidos por Schwartz y Robert M. Lewis en estudios a largo plazo realizados en perros con LES sugieren la intervención de factores genéticos y ambientales en la producción de la enfermedad. Los perros sufrían un síndrome muy parecido al humano, incluyendo las manifestaciones cutáneas y los anticuerpos específicos del suero. Sin embargo, después de una prolongada endogamia, algunos perros mostraron alteraciones séricas sin manifestaciones clínicas de la enfermedad. Esto mismo ocurre en familiares de pacientes con LES. Lawrence E. Shulman, de la Universidad Johns Hopkins, concluye, a partir de una serie de análisis practicados en familiares de pacientes con LES, que hay pruebas claras sobre la existencia de factores genéticos y ambientales.

Si junto a la predisposición genética se requiere un factor desencadenante de las manifestaciones clínicas del LES, ¿de qué factor se trata? Distintas drogas producen un incremento del nivel de anticuerpos antinucleares. Particularmente la procainamida (administrada para tratar la arritmia cardíaca) y la hidralacina (utilizada para el tratamiento de la hipertensión) inducen la formación de anticuerpos contra el ADN de una sola cadena, las nucleoproteínas y, posiblemente, frente a otros antígenos nucleares. La administración prolongada de procainamida determina la aparición de un síndrome similar al LES en alrededor de un tercio de las personas tratadas. Los síntomas cesan al suprimir el tratamiento y raramente se producen alteraciones renales, ni aparecen anticuerpos frente al ADN de doble cadena.

Como modelo experimental, sin embargo, el LES inducido por drogas proporciona datos relativos al mecanismo de inducción de autoanticuerpos. Así, por ejemplo, la procainamida y la hidralacina se metabolizan por acetilación. Marcus M. Reidenberg y Dennis E. Drayer, de Cornell, observaron que los efectos de las drogas eran más pronunciados en personas con un déficit genético del enzima acetilante. Robert G. Lahita, Reidenberg, Drayer y el autor demostraron que aquellos pacientes que recibían, en vez de la forma usual de la procainamida, un análogo acetilado de la misma producían autoanticuerpos más lentamente y en menor cantidad. Hasta ahora carecemos de datos que impliquen a una



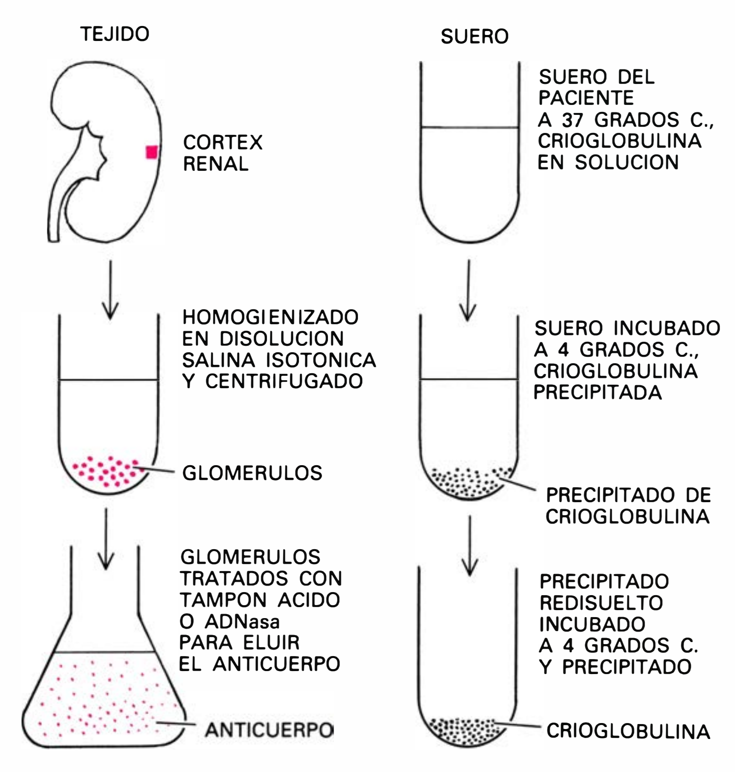
EL TAMAÑO DE LOS COMPLEJOS fue relacionado con la respuesta de anticuerpos y la progresión de la enfermedad en conejos por Frederick G. Germuth, Jr., de la Facultad de Medicina de la Universidad de Johns Hopkins; las propiedades de los complejos se estudiaron en el laboratorio del autor. Para un nivel constante de antígeno inyectado, la patología observada en ocho conejos (*a*) se relacionó con la cantidad de anticuerpo producido en cada animal (*b*). Un test de precipitina (realizado en el laboratorio añadiendo cantidades crecientes de antígeno a una cantidad constante de anticuerpo) muestra cómo la cantidad de complejos inmunitarios precipitados (insolubles) varía con la relación anticuerpo-antígeno (*c*). La curva se representa mediante barras (*d*): la cantidad de precipitado aumenta hasta que el antígeno se encuentra en una cantidad equivalente (*E*) al anticuerpo; luego, disminuye. Los complejos varían en su capacidad para fijar el complemento, el principal determinante de la toxicidad (*e*). El tamaño del complejo, relacionado con la razón anticuerpo-antígeno (*f*), afecta al sitio de depósito (*g*). Los complejos muy grandes (*A*, *B*) tienden a ser eliminados por las células mesangiales y los fagocitos circulantes; los complejos con un peso molecular de alrededor de 500.000 (*C*) se depositan a lo largo de la membrana glomerular y producen una notable alteración de la función renal; los complejos muy pequeños (*D*) recirculan o son filtrados y pasan directamente a la orina sin llegar a lesionar el tejido renal.

droga determinada o agente químico como desencadenante de la enfermedad, pero tal posibilidad debe ser investigada.

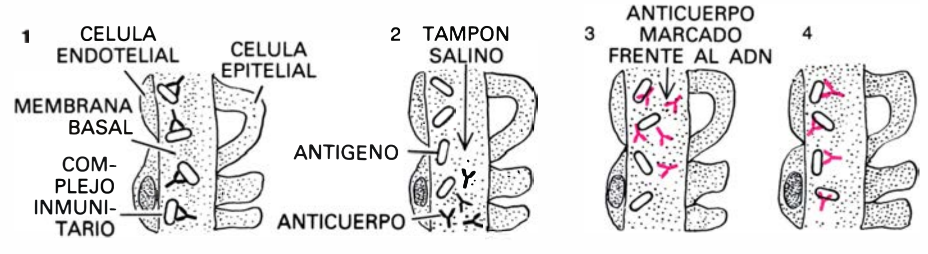
Diversas observaciones sugieren una posible relación con los estrógenos. El LES se da preferentemente en mujeres durante los años de actividad reproductora, cuando el nivel de estrógenos es máximo. Los anticonceptivos orales estrogénicos exacerban a veces la enfermedad. Los hombres sufren raramente el LES, pero se observa con mayor frecuencia en aquellos que padecen una alteración congénita conocida como síndrome de Klinefelter. Tales pacientes poseen dos cromosomas *X* y un cromosoma *Y*; al mismo tiempo manifiestan un metabolismo de los estrógenos anormal. Norman Talal y colaboradores, de la Facultad de Medicina de la Universidad de California en San Francisco, han puesto de manifiesto que la gravedad del LES en ratones se reduce en las hembras desprovistas de estrógenos (por eliminación de las gónadas) o tras la administración de hormonas masculinas. Por otra parte, la enfermedad se agrava en los ratones machos castrados o que han recibido estrógenos. Estudios recientes llevados a cabo por Lahita, Kunkel y Jack Fishman han puesto de manifiesto la existencia de elevados niveles de ciertos metabolitos de los estrógenos en pacientes de ambos sexos con LES.

La sensibilidad a la luz solar que generalmente manifiestan los pacientes con LES hace suponer que la radiación ultravioleta constituye un posible factor de la enfermedad. Las lesiones cutáneas suelen aparecer en regiones de la piel expuestas a la luz del sol. A veces, las lesiones sistémicas se exacerban como consecuencia de la exposición. La irradiación ultravioleta puede alterar las hebras de la doble hélice del ADN. Algunas de las estructuras anormales, como los dímeros de timina (unión de dos bases de los nucleótidos constituyentes del ADN) actúan como determinantes antigénicos en animales de experimentación. La radiación ultravioleta podría incrementar la respuesta inmunitaria y alterar la tolerancia natural al ADN. Tales hipótesis no han sido demostradas, pero pueden considerarse como posibles partiendo de los datos actuales.

El papel de una posible infección vírica en el LES ha sido un tema de considerable especulación e investigación, habiéndose centrado el interés, sobre todo, en los virus de ARN del tipo *C*. Se han detectado proteínas víricas en tejidos de pacientes con LES, aunque el significado de tales hallazgos no está claro. La infección vírica, en lugar de constituir el factor desencadenante, podría ser el resultado de una deficiencia inmunitaria o de la supresión de la respuesta inmunitaria como consecuencia de la administración de drogas para el tratamiento del LES. Material genético de los virus podría incorporarse en las células humanas sin que se produjeran nuevas partículas víricas y sin afectar a la evolución del LES. Una prueba circunstancial sobre la etiología infecciosa de la enfermedad fue presentada por Raphael J. DeHoratius y Ronald P. Messner, que trabajaban en la Facultad de Medicina de la Universidad de Nuevo México. Estos investigadores observaron un aumento de anticuerpos antilinfocitos en



ANTICUERPO FRENTE:	ELUIDO GLOMERULAR	CRIOGLOBULINA
ADN DOBLE CADENA	6 ENTRE 9	21 ENTRE 29
ADN CADENA SIMPLE	8 ENTRE 9	16 ENTRE 28
RIBONUCLEOPROTEINA NUCLEAR	3 ENTRE 9	3 ENTRE 13



LA NATURALEZA DE LOS COMPLEJOS asociados a la enfermedad renal grave se estableció (*a*) examinando muestras de glomérulos obtenidas en autopsias y por análisis de las crioglobulinas, una fracción del suero que contiene complejos inmunitarios y forma precipitados geliformes a bajas temperaturas. Ciertos anticuerpos estaban presentes en los glomérulos y en las crioglobulinas en concentraciones muy superiores a las observadas en el suero, lo que indicaba que tendían a formar complejos inmunitarios. La tabla muestra, para cada uno de los tres anticuerpos, la proporción de casos en los que la concentración preferencial excedía las 50 veces. Otra prueba de la localización de los complejos inmunitarios en los glomérulos se obtuvo (*b*) disociando el anticuerpo de los complejos en secciones de riñón, añadiendo anticuerpo anti-ADN marcado y observando que la unión al antígeno ADN tenía lugar en los mismos sitios que los depósitos de anticuerpo. La ilustración es de Bunji Tagawa.

personas que convivían con pacientes de LES y que no eran familiares. Los estudios llevados a cabo por Schur, Kunkel y el autor y por Talal sobre anticuerpos que reaccionan con el ARN han suscitado la posibilidad de que estén implicados antígenos víricos. No obstante, resulta difícil descartar la intervención de antígenos celulares.

Actualmente conocemos mucho mejor los mecanismos que intervienen en la producción de lesiones tisulares que la etiología de la enfermedad. Todo lo más que puede decirse sobre la causa del LES es que se trata probablemente de la interacción de factores genéticos y otros factores tales como drogas, hormonas y virus. Los avances de la investigación sobre la etiología del LES están condicionados por nuevos descubrimientos sobre los mecanismos implicados en las funciones inmunitarias normales.

Las posibilidades cada vez mayores de realizar un diagnóstico precoz de la enfermedad, junto a la clarificación de los mecanismos determinantes de las lesiones tisulares y a la mayor facilidad para detectar los períodos activos de la enfermedad mediante análisis de sangre, han mejorado el tratamiento sintomático del LES y particularmente de las manifestaciones renales. Actualmente, más del 90 por ciento de los pacientes sobreviven cinco años después del diagnóstico de la enfermedad, mientras hace 25 años sólo sobrevivían entre el 50 y 60 por ciento. Al aumentar las posibilidades para un mejor diagnóstico se ha puesto de manifiesto una incidencia de la enfermedad mayor de la esperada.

Los estudios llevados a cabo sobre la inmunología del LES han contribuido a esclarecer los mecanismos básicos a través de los cuales determinados agentes inmunitarios provocan lesiones tisulares. Evidentemente, algo anormal ocurre en el ecosistema inmunológico; la frágil simbiosis de sus diversos elementos (células *B*, células *T*, anticuerpos y componentes del complemento, entre otros) ha resultado alterada. La investigación de los trastornos funcionales del sistema inmunitario en el LES permitirá interpretar el "lenguaje" mediante el cual los diversos elementos se comunican entre sí. Por otra parte, contribuirá sin duda a un mejor conocimiento de las alteraciones inmunológicas propias de otras enfermedades tales como la artritis reumatoide, la esclerosis múltiple y la glomerulonefritis crónica, entre otras.

Astronomía de líneas de rayos gamma

Los procesos astrofísicos emiten radiación electromagnética en bandas muy estrechas a energías superiores incluso a las de los rayos X. Las líneas espectrales resultantes se pueden detectar con instrumentos montados en globos o satélites

Marvin Leventhal y Crawford J. MacCallum

Lo que se sabe acerca de la naturaleza del universo se ha deducido casi enteramente a partir de los fotones, o cuantos de radiación electromagnética, que alcanzan la proximidad de la Tierra. Hasta hace medio siglo, los astrónomos solamente podían detectar fotones de energías entre 1,5 y 3,5 electronvolt: los fotones de la luz visible. Después empezaron a ampliar el dominio de energías de los fotones hacia valores más bajos, en ciertas porciones de las regiones infrarroja y de radio del espectro electromagnético y hacia valores más altos en el ultravioleta próximo. Con la llegada de los cohetes, globos para gran altura y satélites artificiales, estuvieron en condiciones de ampliarlo hacia valores mucho más altos, hasta el dominio de energía de fotones que no pueden penetrar en la atmósfera terrestre: los fotones del ultravioleta extremo, rayos X y rayos gamma.

Los rayos gamma constituyen la forma de mayor energía de la radiación electromagnética; la energía de sus fotones se mide en millones de electronvolt (MeV), y en principio no tiene límite superior. Los fotones de rayos gamma procedentes del espacio se detectaron por primera vez hace un par de decenios. Los primitivos detectores simplemente registraban la llegada de los fotones, sin ser capaces de analizar sus energías de la manera como los fotones de luz visible se resuelven en líneas de distinta energía por medio de un espectrógrafo. En la actualidad, sin embargo, se han desarrollado instrumentos que pueden detectar líneas espectrales de rayos gamma. Están comenzando a proporcionar información sobre procesos de alta energía y objetos que llaman la atención de los astrónomos modernos, tales como las supernovas, estrellas de neutrones y fenómenos en el centro de las galaxias. Mientras que las líneas del espectro óptico se producen como efecto de transiciones entre los niveles energéticos de los electrones en los átomos, las líneas en el espectro de rayos gamma son consecuencia de transiciones entre los niveles de energía de los núcleos atómicos. Considérese el núcleo del neón 22, que consta de 10 protones y 12 neutrones. El núcleo está generalmente en su estado fundamental, o nivel de mínima energía, pero puede también ocupar niveles excitados de energía. Si el núcleo del neón 22 está en el más bajo de los niveles excitados de energía, saltará al nivel básico en cuestión de unas billonésimas de segundo. La transición irá acompañada de la emisión de un fotón de rayos gamma con una energía de precisamente 1,2746 MeV. Esa cantidad de energía corresponde a la diferencia entre los dos estados. El rayo gamma es por tanto el testimonio de una transición nuclear específica.

La transición puede producirse como consecuencia de varios sucesos diferentes. En uno de ellos, un núcleo de sodio 22, que es radiactivo, baja a un estado excitado de neón 22 con una semivida de 2,6 años. Se cree que tal transición es corriente después de la formación de una nova. En otro suceso, una partícula subatómica de alta energía choca con un núcleo de neón 22 en su estado fundamental y lo excita a un nivel de mayor energía. En un tercer suceso, la captura de un neutrón por un núcleo de neón 21 da como resultado un núcleo excitado de neón 22. En cada caso, la subsiguiente –en realidad, la casi inmediata– transición del núcleo desde el estado excitado hasta el fundamental va acompañada de la emisión de un fotón de rayos gamma de energía característica.

Las líneas de rayos gamma pueden también resultar de procesos en los que no intervienen núcleos atómicos. Uno de estos es el choque de un electrón y un positrón, o antielectrón. Los dos tienen idénticas propiedades excepto en su carga eléctrica, que es negativa para el electrón y positiva para el positrón. Cuando se encuentran un electrón y un positrón, se aniquilan mutuamente. Su masa se libera en forma de energía electromagnética, radiada desde el punto de aniquilamiento y generalmente en la forma de dos fotones gamma, cada uno con una energía de 0,511 MeV. El aniquilamiento electrón-positrón parece ser un proceso particularmente corriente cerca de las estrellas de neutrones y en el centro de nuestra galaxia.

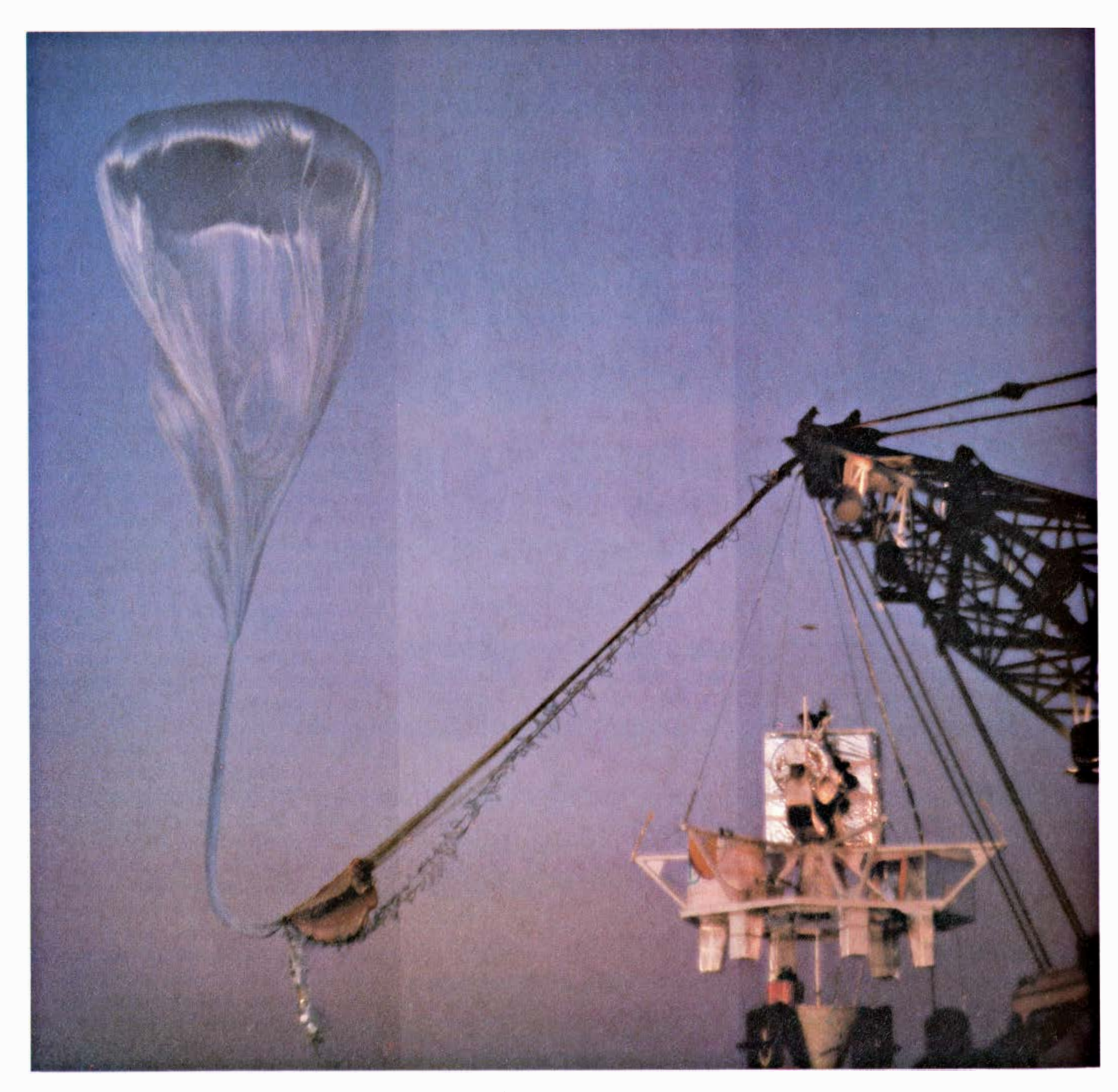
Un segundo proceso no nuclear que da lugar a la emisión de líneas de rayos gamma es la emisión por electrones de ciclotrón. Los electrones que se mueven en un campo magnético están sometidos a una aceleración circular que los hace radiar fotones de energía característica. La energía es proporcional a la intensidad del campo, y si el campo es suficientemente intenso y uniforme, como puede ser en los polos magnéticos de una estrella de neutrones, los electrones radian rayos gamma.

Las primeras líneas de rayos gamma inequívocamente detectadas como procedentes del espacio fueron producidas por varios de estos mecanismos. Emanaron de una fulguración solar gigante que se produjo en agosto de 1972, y fueron detectadas por Edward L. Chupp y sus colaboradores, de la Universidad de New Hampshire, que había colocado un pequeño detector de rayos gamma a bordo del satélite *OSO-7* (el séptimo Observatorio Solar Orbital de

la Administración Nacional de Aeronáutica y del Espacio, NASA). Las líneas incluían una a 0,511 MeV, originada en el aniquilamiento en la atmósfera solar de positrones producidos en la fulguración solar. Otra línea, a 2,2 MeV, se originó por la captura de neutrones por protones y la emisión de un rayo gamma por cada uno de los núcleos de deuterio resultantes. Finalmente, dos débiles líneas a 4,4 y 6,1 MeV se originaron por la transición a sus estados fundamentales de núcleos de carbono 12 y oxígeno 16 que habían sido llevados a estados energéticos excitados en la atmósfera solar por choques con protones acelerados en la fulguración.

En los ocho años que han transcurrido desde que se hicieron esas primeras observaciones, varios grupos de investigación, incluyendo el nuestro de los Laboratorios Bell y los Laboratorios Sandia, han estado realizando experimentos con telescopios de rayos gamma transportados en su mayoría por globos, pero algunas veces por satélites, en intentos de elevar el instrumento por encima de la casi totalidad de la atmósfera terrestre y detectar líneas de rayos gamma de origen astrofísico. El campo está todavía en su infancia. Aquí pasaremos revista a algunas de sus perspectivas, describiremos un telescopio de líneas de rayos gamma y daremos algunos de los primeros resultados.

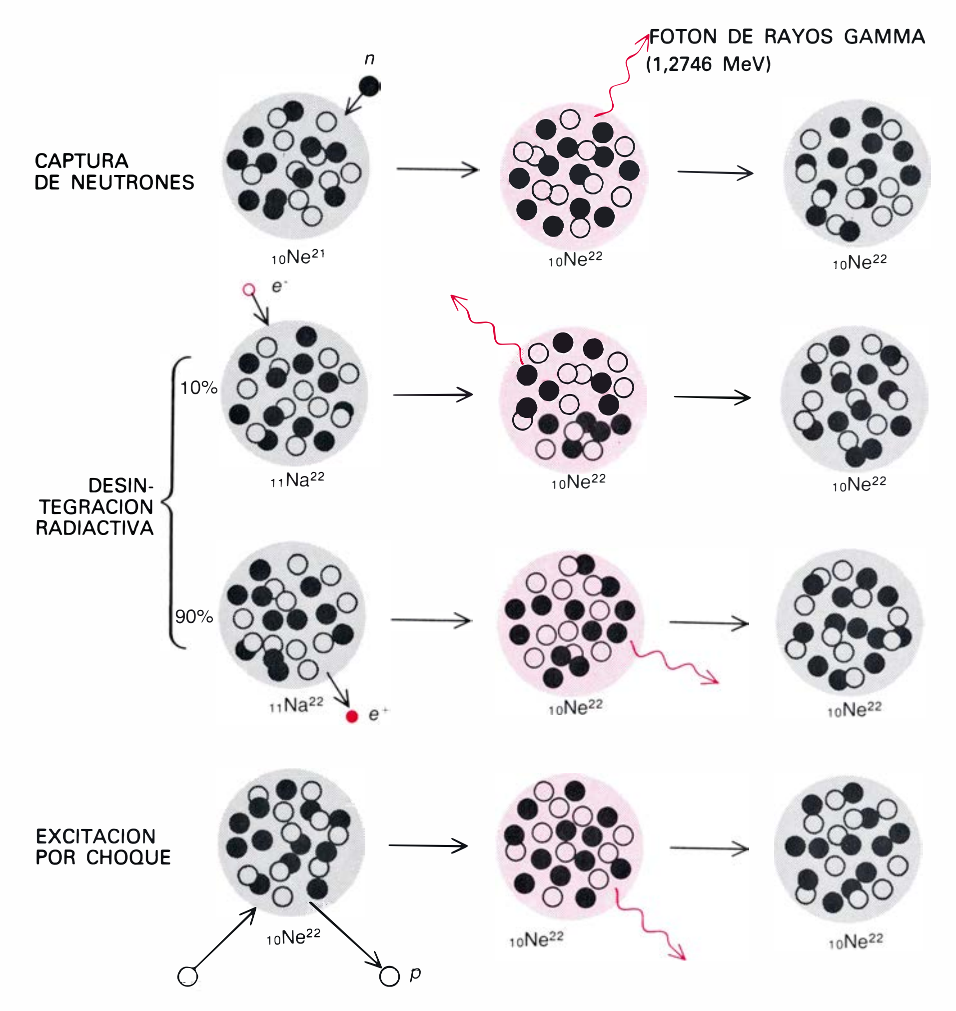
Quizá la más importante posibilidad es la de hacer observaciones que tienen relación con la síntesis cósmica de los elementos químicos. Se supone que la gran explosión creó principalmente hi-



GLOBO LANZADO en la madrugada del 9 de octubre de 1975 para elevar el telescopio de líneas de rayos gamma de los autores en un vuelo de prueba sobre Alamogordo, Nuevo México. El globo está unido a un paracaídas, conectado a la grúa de la derecha, del que se suspende la carga útil. La base de esta carga soporta un grupo de almohadones antichoque. El cono subyacente a estos almohadones es lastre. La caja de la parte superior de la carga contiene la instrumentación electrónica. En la parte anterior de la caja hay un bidón de nitrógeno líquido usado como refrigerante. El telescopio se halla entre ambos y por ello no se puede ver. Al final del vuelo, el paracaídas con la carga se separa del globo y abre éste, que también regresa a tierra.

drógeno y helio, que se convierten en elementos más pesados por fusión nuclear en el interior de las estrellas. Sin embargo, para núcleos más pesados que el hierro 56, la fusión no libera energía. Por consiguiente, hay que invocar en las estrellas un mecanismo distinto de la fusión para dar cuenta de la creación de estos elementos. Específicamente, se supone que los elementos intermedios y los más pesados se producen en su mayoría en la explosión de estrellas.

Los detalles se indican en lo que sigue. Una estrella es una esfera de gas en que el calor producido por la fusión nuclear contrarresta la tendencia del gas a caer hacia dentro. Con el tiempo, sin embargo, la estrella agota su provisión de combustible nuclear. Una hipótesis ampliamente aceptada mantiene que si la masa de la estrella es comparable a la masa del Sol, la estrella se contrae gradualmente hasta que tiene aproximadamente el radio de la Tierra. Se ha convertido ahora en una enana blanca; ha agotado su combustible nuclear y se está enfriando. Si la masa es substancialmente más grande, la estrella no se contrae gradualmente sino que colapsa catastróficamente y explota. Se estima que una vez cada 10 a 300 años, una estrella de entre los cientos de miles de millones que existen en una galaxia típica alcanza este punto final de la evolución estelar, y la luz de la explosión puede ser más brillante que la totalidad de la galaxia durante un período de días o incluso semanas. Este es el suceso que se llama supernova. Deja una estela de residuos dispersos en expansión, de los que el mejor ejemplo en nuestra galaxia es la nebulosa del Cangrejo en Tauro.

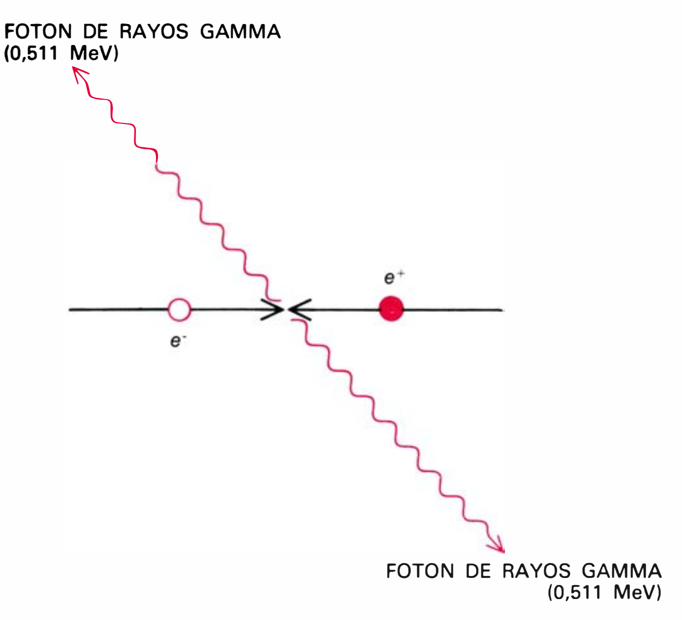


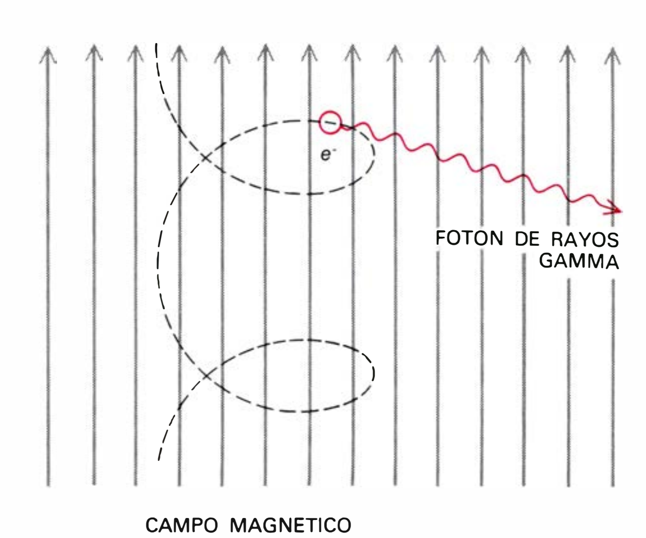
EMISION DE UN RAYO GAMMA POR UN NUCLEO ATOMICO al descender de un nivel excitado de energía a su estado fundamental, o nivel de mínima energía posible. Aquí se usa el núcleo de neón 22 para esquematizar tres maneras en las que puede producirse la emisión. En la captura de neutrones (*fila superior*) el neón 21 se convierte en neón 22 en un estado excitado por la adición de un neutrón (*n*). En la desintegración radiactiva (*segunda y tercera filas*), el sodio 22 se transmuta en neón 22 en un estado excitado por la adición de un electrón (e^-), generalmente capturado de su órbita en el átomo, o la emisión de un positrón (e^+), que es la antipartícula del electrón. La última posibilidad es la más corriente de las dos, pero cualquiera de ellas convierte un protón en un neutrón. En la excitación por choque (*última fila*), la energía impartida al neón 22 por una partícula energética (aquí un protón, *p*) lo eleva a un estado excitado. En cada caso, el retorno del núcleo excitado a su nivel fundamental libera un fotón (cuanto de radiación electromagnética) con una energía característica de 1,2746 millones de electronvolt (MeV).

Los astrofísicos teóricos W. David Arnett, Donald D. Clayton y Stirling A. Colgate fueron de los primeros que desarrollaron modelos de ordenador aplicables a las reacciones nucleares que tienen lugar en esos pocos segundos explosivos, cuando el proceso de la nucleosíntesis estelar, que hasta entonces puede haber durado miles de millones de años, toma caminos enteramente nuevos. Tales modelos sugieren que la explosión tiende a construir núcleos más pesados a partir de otros más ligeros. En particular, parece posible que las supernovas hayan producido la mayor parte de los elementos intermedios y pesados en las abundancias cósmicas relativas que se observan.

Por supuesto, sería buena idea someter esas teorías a prueba, y en principio esto se puede hacer. Las rayas espectrales para energías correspondientes a la luz y a los rayos X revelan cuáles son los átomos presentes en los restos de una explosión de supernova. Sin embargo, los átomos que emiten la radiación pueden haber estado formando parte de la estrella ya antes de la explosión, e incluso algunos de ellos pueden haber sido aportados desde el espacio interestelar. Por otra parte, muchos de los núcleos sintetizados en la explosión deben ser radiactivos y la desintegración de tales núcleos va acompañada por la emisión de rayos gamma a energías características. (Algunas de las energías características de emisión son ligeramente más bajas y corresponden a rayos X de alta energía.) La detección incluso de una sola línea de rayos gamma que procediera del lugar donde se ha producido una explosión de supernova representaría un avance de gran importancia en nuestra comprensión del modo como se edificó nuestro actual universo.

Un segundo tipo de explosión estelar es la nova. Las novas son mucho menos violentas que las supernovas, pero mucho más frecuentes; se producen quizá 25 veces por año y por galaxia. También aportan una destacada contribución a la nucleosíntesis. En la hipótesis actualmente aceptada, una explosión de nova se produce en un sistema estelar binario en el que una estrella es ordinaria y la otra una enana blanca, estrella que, como hemos visto, ha agotado su combustible nuclear sin explotar. Aunque una enana blanca no tiene fuente propia de energía, se cree que, cuando tal estrella es miembro de un sistema binario, va acumulando lentamente material, principalmente





DOS PROCESOS NO NUCLEARES que generan rayos gamma. En el aniquilamiento electrón-positrón (*izquierda*), un positrón encuentra a un electrón, su antipartícula, y la masa de ambas se convierte en energía electromagnética que toma la forma de dos fotones de rayos gamma, cada uno portador de 0,511 megaelectronvolt MeV. En la emisión por electrones de ciclotrón (*derecha*) un campo magnético intenso y uniforme (representado por líneas de campo densas y paralelas) añade un movimiento circular a la trayectoria de un electrón. La frecuencia de los bucles corresponde a la frecuencia (y por tanto a la energía) de cada fotón que el electrón emite. Si la frecuencia es suficientemente alta, los fotones serán rayos gamma.

hidrógeno, atraído de su estrella compañera. Con el tiempo, la temperatura y presión en la base de esta capa gaseosa acumulada alcanzan valores explosivos, y la envoltura estalla.

Los cálculos de nucleosíntesis realizados para el caso de explosiones de nova por grupos dirigidos por Clayton y James W. Truran, Jr., sugieren una gran producción de sodio 22, casi todo el cual se desintegra en forma del estado excitado de neón 22 con una vida media de 2,6 años. Varios meses después de la explosión, la línea de rayos gamma más prominente de las procedentes del lugar de la explosión debería ser, por tanto, una línea a 1,2746 MeV, que representa el descenso de los núcleos de neón recién sintetizados a su estado fundamental. Varios meses debe ser un tiempo suficiente para que los astrónomos puedan elevar en un globo un telescopio de rayos gamma. Como las novas son frecuentes, y es muy probable que una de ellas tenga lugar cerca del sistema solar, la línea de 1,2746 MeV podría muy bien ser la primera línea de rayos gamma detectada como procedente de una estrella en explosión.

Las mismas explosiones de supernova, que pueden contribuir a edificar los elementos intermedios y pesados, dejan en el lugar de la explosión un resto estelar que tiene aproximadamente la masa del Sol comprimida en un radio de unos 10 kilómetros. Se calcula que la densidad en el centro del resto es de unos 10^{15} gramos por centímetro cúbico, valor que excede a la densidad en el interior del núcleo atómico. En realidad, la densidad es tan grande que se cree que los electrones del núcleo han sido aplastados contra los protones anulando la carga positiva de éstos y convirtiéndolos en neutrones. El resto se conoce por ello como estrella de neutrones.

La juventud de un tal objeto es diferente de su edad avanzada, por lo que se sabe de una y otra mediante deducciones realizadas por los astrofísicos en la Tierra. Las estrellas de neutrones jóvenes y en rápida rotación son radio-pulsars: fuentes de radiación electromagnética pulsante a frecuencias de radio, detectadas por primera vez en el decenio de 1960. La fuente de energía para las emisiones de los pulsars parece ser una constante pérdida en la energía cinética de rotación de la estrella de neutrones, porque se ha observado que, en todos los pulsars conocidos, el período entre impulsos sucesivos decrece lentamente.

El mecanismo real mediante el cual la energía de rotación se convierte en la radiación observada continúa siendo un misterio. Hay muchas teorías contradictorias. En general, sin embargo, se postula que la estrella de neutrones tiene un campo magnético encerrado en ella, porque se admite que el campo de la estrella padre se conservó a través de los diversos sucesos que crearon el resto estelar. Como dicho resto está comprimido, también lo está el campo magnético, que por ello alcanza una enorme intensidad: se estima valer 10^{13} gauss. (La intensidad del campo magnético terrestre es una fracción de gauss.) La estrella residual está en giro, de modo que también lo está el campo magnético. El resultado es que el campo magnético induce un campo eléctrico en la superficie de la estrella. El campo eléctrico atrae partículas cargadas de la superficie y las acelera hasta altas energías. Las trayectorias de las partículas se curvan entonces por efecto del campo magnético. El encorvamiento de las trayectorias hace que las partículas radien fotones.

El carácter pulsante de la radiación observado desde la Tierra se explica suponiendo que los dos polos del campo magnético no están alineados con el eje de rotación de la estrella. De aquí que, a medida que la estrella de neutrones gira, cada polo del campo, en el cual la intensidad de éste es máxima, aparece y desaparece alternativamente respecto a cualquier visual. Por eso, la radiación que procede de ella se detecta en forma de destellos, lo que se llama efecto de faro.

¿Por qué razón tiene todo esto importancia en relación con la astronomía de líneas de rayos gamma? El primero en indicar una razón fue Peter A. Sturrock, de la Universidad de Stanford. En la mayoría de los modelos teóricos aceptados, el campo magnético de la estrella de neutrones genera grandes flujos de positrones. Estos se producen de una manera curiosa. El campo magnético es tan intenso que, cuando incurva la trayectoria de una partícula cargada de alta energía, la

partícula emite fotones de rayos gamma. Si la energía de un tal fotón es superior a 1,022 MeV, cuando el fotón encuentra un campo magnético puede dar lugar a una pareja electrón-positrón. (La energía del fotón debe ser superior a 1,022 MeV porque la masa de un electrón o un positrón corresponde a 0,511 MeV por aplicación de la fórmula $E = mc^2$.) Cuando un positrón creado así encuentra a otro electrón, la pareja a su vez se aniquila, dando lugar a fotones de rayos gamma que producen una línea a 0,511 MeV. La detección de esta radiación de frecuencia fija debería permitir a los astrónomos acotar la producción de positrones por los pulsars, y en consecuencia, aplicar una importante ligadura a los modelos de pulsars que pueden permitirse.

En aproximadamente un millón de años, hasta la estrella de neutrones en giro más rápido empieza a retardarlo. Su vida como radio-pulsar se acaba, pero su vejez puede resultar todavía de interés. En nuestra galaxia se han descubierto cientos de fuentes de rayos X aparentemente puntuales. Se cree que la mayor parte son antiguas estrellas de neutrones que están acumulando materia procedente de estrellas compañeras en un sistema binario. A medida que la materia acumulada se arremolina en el campo gravitatorio de la estrella de neutrones, se acelera hasta altas energías y forma un disco ardiente de acumulación. Los choques entre las partículas en el disco generan el ancho espectro de rayos X que se observa desde la Tierra.

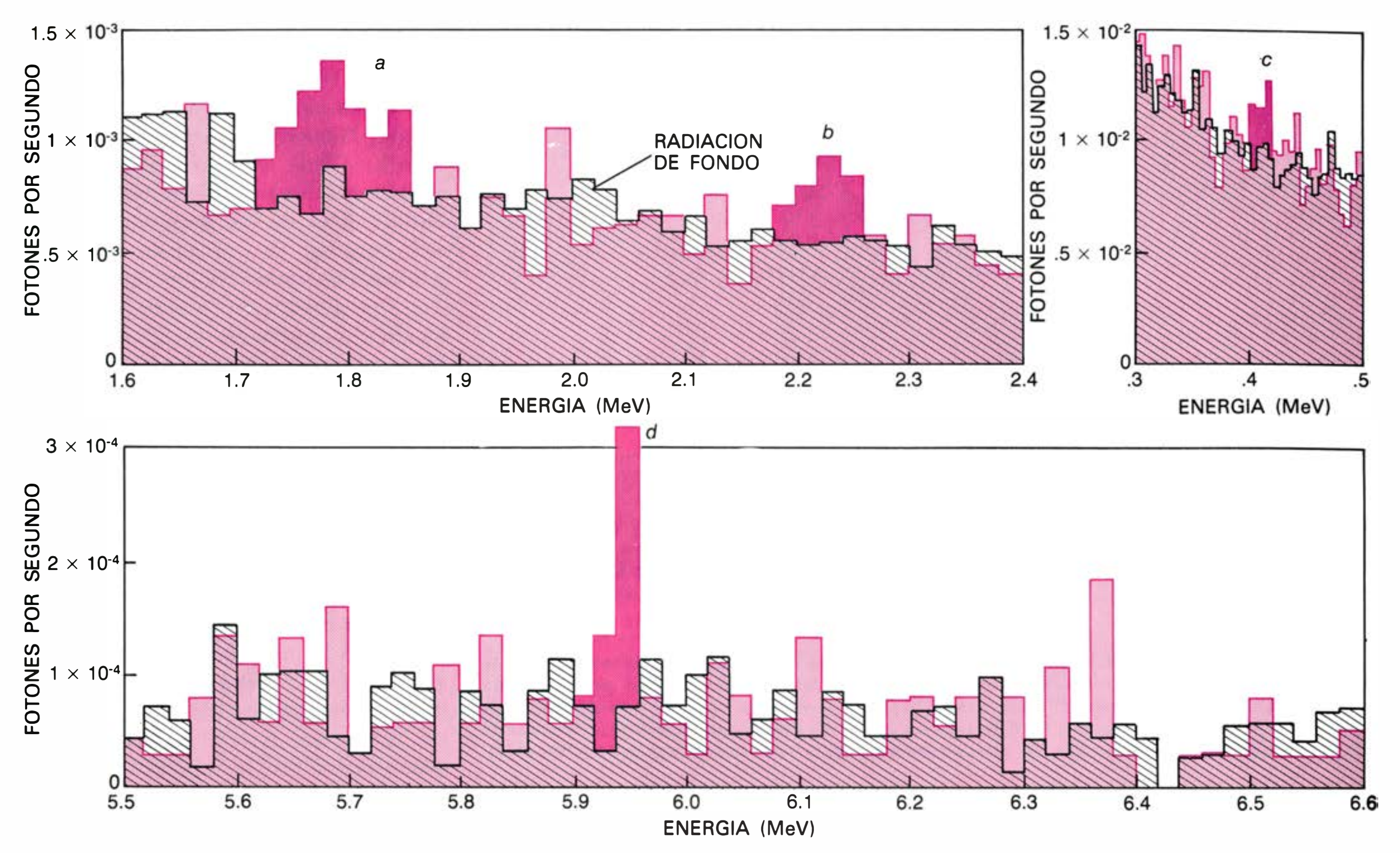
Se asume que la materia que se acumula está altamente ionizada y, por tanto, deberá constar principalmente de protones y electrones de alta energía. Y habría también la posibilidad de detectar líneas de rayos gamma, porque los protones de alta energía que inciden sobre la superficie de la estrella de neutrones excitarían a sus núcleos y el retorno de éstos a su estado fundamental generaría fotones de rayos gamma. La naturaleza de la superficie de una antigua estrella de neutrones se podría deducir a partir de las energías características de los rayos gamma emitidos.

Una piedra angular bien establecida de la teoría de la relatividad sostiene que un fotón pierde energía al moverse en contra de un campo gravitatorio, y que su longitud de onda aumenta correspondientemente; en el caso de la luz, el fotón tiende a hacerse más rojo. El cambio es proporcional a la variación de la energía potencial gravitatoria, la cual, para el escape desde la superficie de una estrella, es proporcional a la masa de la estrella dividida por su radio. Por ser grande la masa de una estrella de neutrones y pequeño su radio, su energía potencial gravitatoria resulta ser enorme y también lo es la pérdida de energía de los fotones que escapan de su superficie. Hay que esperar, por tanto, que todas las líneas de rayos gamma se corran hacia energías menores (y longitudes de onda mayores) en cuantía considerable. El parámetro de corrimiento hacia el rojo en la superficie se designa por Z_s. Para una línea dada cualquiera, se detectaría la diferencia entre la energía observada por fotón y la energía por fotón si la radiación hubiese sido emitida en la Tierra, dividida por la energía observada por fotón.

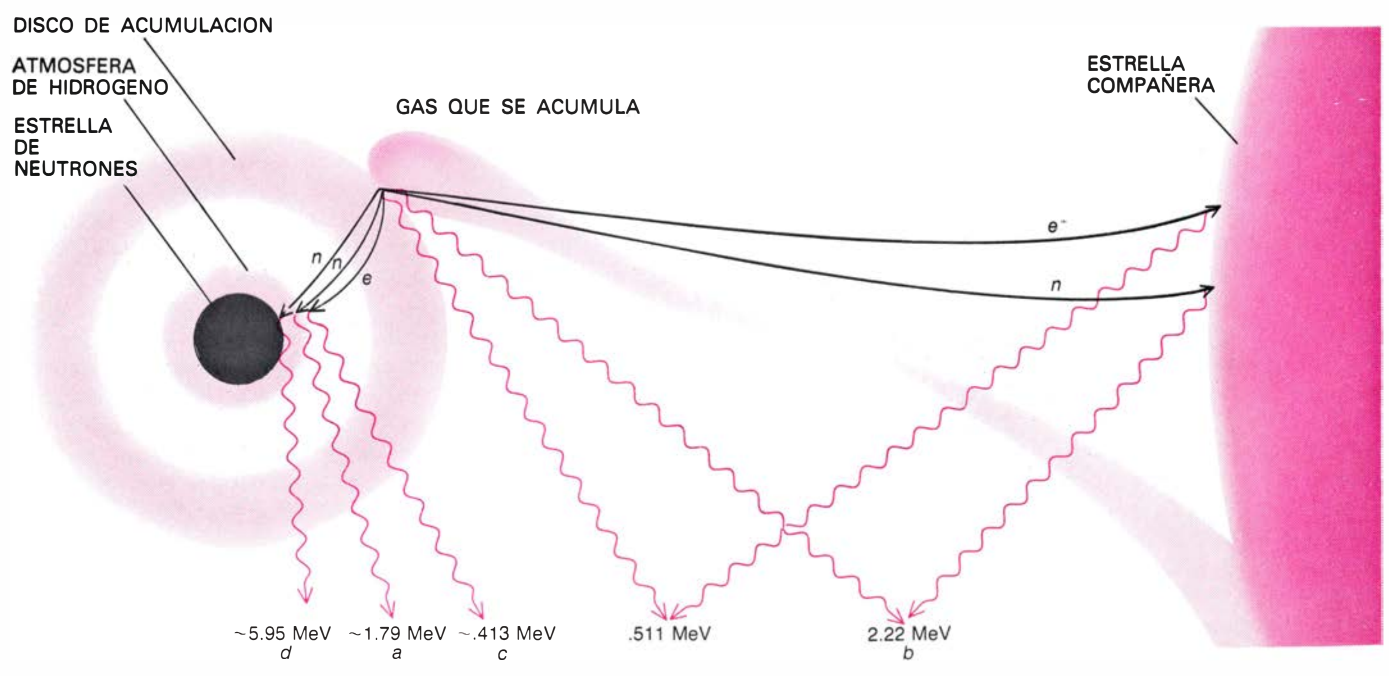
Aparece ahora una segunda línea de argumentación, desarrollada independientemente por Kenneth Brecher y por Richard Bowers. Varias estrellas de neutrones conocidas tienen tales órbitas en su sistema binario que son eclipsadas periódicamente por su estrella compañera. Para cada una de esas estrellas de neutrones, las interrupciones en la radiación X detectada en la Tierra revelan el período orbital y el tiempo de eclipse. Mientras tanto, el movimiento orbital de la estrella compañera las acerca y aleja alternativamente de la Tierra, de modo que su emisión luminosa es alterada por efecto Doppler. Además, la intensidad de la luz está modulada porque la estrella compañera cambia de forma al responder a las mareas en el campo gravitatorio de la estrella de neutrones. Reunidas, todas esas observaciones permiten obtener los parámetros orbitales de la estrella de neutrones, y cuando se conoce la órbita, puede calcularse la masa de la estrella.

El punto siguiente que puede hacer variar la argumentación es la posibilidad de que la estrella de neutrones esté en realidad constituida en gran

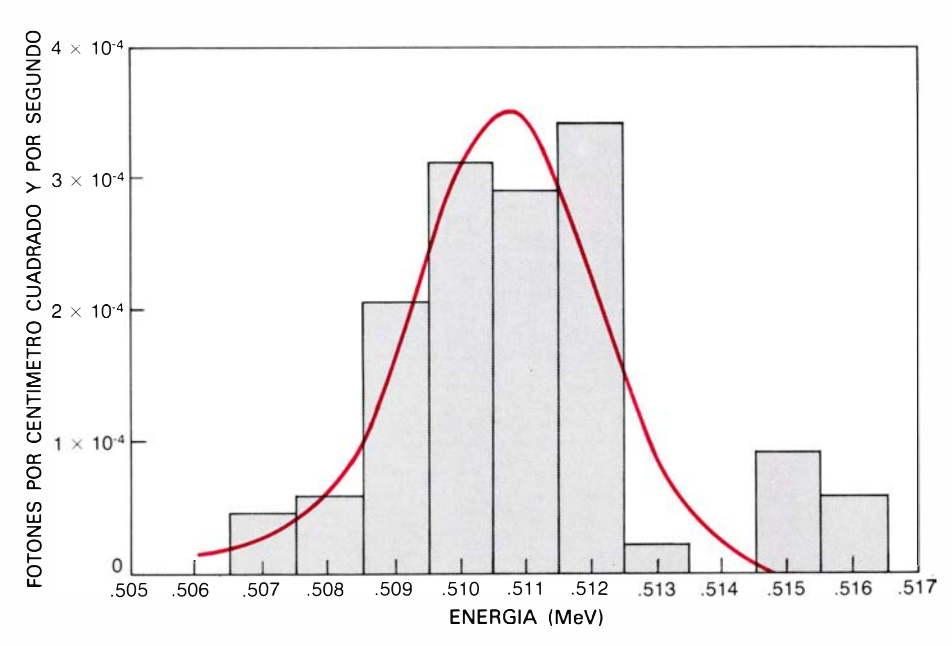
ESPECTRO DE RAYOS GAMMA DE HERCULES X-1, prominente fuente de rayos X. Incluye dos máximos, o características similares a líneas, de rayos gamma a 55 keV y 110 keV. La línea a 55 keV se interpreta como la radiación de ciclotrón emitida por electrones en el intenso campo magnético de una estrella de neutrones, o resto superdenso de una explosión estelar. La línea a 110 keV se interpreta como un armónico a una energía doble de la de emisión de ciclotrón. Los datos fueron registrados el 3 de mayo de 1976 por un globo instrumentado lanzado cerca de Palestina por un grupo dirigido por Trümper. Se presume que el resto del espectro, incluyendo los datos a la izquierda (*línea de trazos*), obtenido por un instrumento montado en un satélite, es el final de la parte alta de un espectro continuo de rayos X emitidos por átomos que chocan en un disco de materia que rodea a la estrella de neutrones.



BROTE O EXPLOSION DE RAYOS GAMMA procedente de una dirección en el cielo opuesta a la del centro de nuestra galaxia, que se produjo el 10 de junio de 1974 y duró 20 minutos. Aquí se presenta su espectro en tres diferentes bandas de energía. Cuando el telescopio apuntó a esa parte del cielo, cuatro características semejantes a líneas destacaron marcadamente por encima de la radiación de fondo. Dos de ellas, señaladas *a* y *b*, pueden representar la creación de fotones de rayos gamma por la reacción que liga un neutrón y un protón para formar un núcleo de deuterio. La reducción del rasgo *a* a una energía inferior a la de *b* se explica suponiendo que los fotones que contribuyen a *a* se produjeron cerca de la superficie de una estrella de neutrones y perdieron energía al escapar del potente campo gravitatorio de la estrella. La característica *c* se atribuye a aniquilación electrón-positrón, siendo la energía de los fotones también inferior a su valor normal de 0,511 MeV. La característica *d* se atribuye a la captura de neutrones por el núcleo del hierro 56, convirtiendo cada uno de esos núcleos en un estado excitado del hierro 57, que desciende a su estado fundamental mediante la emisión de un fotón. El hierro 56 es el núcleo atómico más estable, y por ello en muchos modelos teóricos forma la superficie de una estrella de neutrones. De nuevo la energía de los fotones es inferior a la normal. Datos recogidos por el grupo de Allan S. Jacobson en el Laboratorio de Propulsión del Cal Tech.



ACUMULACION EPISODICA DE GAS sobre una estrella de neutrones procedente de su compañera en un sistema binario de estrellas. Ese proceso pudo haber causado el brote de rayos gamma cuyo espectro se muestra en la parte superior. En esta hipótesis, los choques de gas incidente con el gas del disco que rodea la estrella de neutrones pueden originar neutrones y positrones. Los acontecimientos subsiguientes, incluidos la captura de neutrones por hidrógeno y por hierro 56 y la aniquilación mutua de electrones y positrones, crean entonces los fotones de que consta el brote. Los fotones emitidos en, y cerca de, la superficie de la estrella de neutrones pierden energía cuando escapan del campo gravitatorio de la estrella. Las letras de *a* a *d* identifican las líneas de rayos gamma que aparecen en las figuras de arriba. La ilustración se inspira en otra debida a Lingenfelter, Higdon y Ramaty.



RADIACION GAMMA PROCEDENTE DEL CENTRO GALACTICO (o al menos de una región en la dirección del centro galáctico). Esta radiación abarca una especie de línea a 0,511 MeV cuya intensidad corresponde a la aniquilación de 10^{43} positrones por segundo. Se sigue sin conocer el mecanismo a través del cual se producen los positrones. La figura resume los datos recogidos por el telescopio de líneas de rayos gamma de los autores durante un vuelo con globo por encima de Alice Springs (noviembre, 1977).

parte por partículas subatómicas más exóticas que los neutrones. En la literatura científica se han detallado muchas posibilidades diferentes, entre las que se ha llegado a incluir a los mismos quarks; varios físicos teóricos han indicado que diferentes composiciones conducen a diferentes radios (R) para una masa (M) dada. Para cada estrella de neutrones, la determinación de M y la medida de Z_s, que es proporcional a M/R, da por resultado la determinación de R. Esto limitará considerablemente las propuestas referentes a la composición de la estrella. Por tanto, la astronomía de líneas de rayos gamma conduce directamente a la física de la materia superdensa.

En relación con ello, puede ser importante observar que se ha identificado recientemente una nueva clase de objetos astronómicos. Los objetos emiten enormes flujos de radiación gamma durante períodos de unos segundos o minutos, y después la emisión cesa. Todavía no se sabe mucho acerca de esos objetos. La especulación más probable es la que sostiene que las explosiones o brotes de rayos gamma se producen por sucesos energéticos en, o cerca de, la superficie de las estrellas de neutrones. Los brotes podrían tener lugar, por ejemplo, si la acumulación de materia sobre una estrella de neutrones, procedente de su estrella compañera en un sistema binario, se presentara episódicamente.

El 10 de junio de 1974, un grupo dirigido por Allan S. Jacobson, del Laboratorio de Propulsión a Chorro del Instituto de Tecnología de California, ordenó a un telescopio de líneas de rayos gamma y gran ángulo de visión, montado en un globo de gran altitud, que apuntase en la dirección opuesta al centro de la galaxia, y detectó un brote de rayos gamma que duró 20 minutos. Para el grupo, el brote era de origen astrofísico. En el espectro del brote parecían apreciarse algunos detalles semejantes a líneas. Se ha propuesto que representaban varios mecanismos comunes que generan radiación gamma, con la peculiaridad de que estas líneas manifestaban un corrimiento hacia el rojo de la misma proporción que el que aparece para energías más bajas. Así, un rasgo semejante a una línea a 0,4 MeV en el espectro del brote se interpretó como la línea a 0,511 MeV correspondiente a la aniquilación positrón-electrón, que ha sido corrida hacia el rojo en un Z de 0,28. Sería en efecto difícil encontrar en tal región del espectro electromagnético otra transición nuclear común que pudiera explicar tal línea. Por otro lado, un valor de 0,28 para el parámetro de corrimiento hacia el rojo en la superficie de una estrella de neutrones concuerda bien con los cálculos teóricos.

Una observación independiente realizada por nuestro propio grupo reveló una posible línea a 0,4 MeV cuando se apuntó nuestro instrumento, transportado por un globo, en la misma dirección. El campo de visión del instrumento era grande, pues incluía la nebulosa del Cangrejo y la estrella de neutrones en su centro. Más recientemente, Bonnard J. Teegarden y Thomas L. Cline, del Centro de Vuelos Espaciales Goddard de la NASA, han dado cuenta de las primeras medidas del espectro de un brote de rayos gamma obtenido con un instrumento de alta resolución montado en un satélite. También ellos encontraron indicaciones de corrimiento hacia el rojo de las líneas de rayos gamma inducido por la gravedad. Parece muy posible que las primeras medidas de corrimiento hacia el rojo producido por la gravedad en la radiación de una estrella de neutrones estén al alcance de la mano.

Los electrones que inciden sobre la superficie de una estrella de neutrones presentan la posibilidad de una observación más, porque tienen un movimiento circular de ciclotrón en el campo magnético de la estrella, y por consiguiente radian fotones. Donde el campo es grande y uniforme, cerca de los dos polos de la estrella, la emisión de ciclotrón podría consistir en rayos gamma. Es una posibilidad propuesta en primer lugar por el astrofísico ruso Rashid A. Sunyaev y sus colegas. La detección de tales líneas equivaldría a la medida de la intensidad del campo. Sería una prueba de que los enormes campos magnéticos están allí.

En mayo de 1976, un grupo del Instituto Max Planck de Física y Astrofísica en Garching, Alemania Occidental, trabajando en colaboración con un grupo de la Universidad de Tübingen, y ambos equipos a las órdenes de Joachim Trümper, lanzó un globo de gran altitud sobre Palestina, Texas. Llevaba un instrumento con el que obtuvieron un espectro de la radiación gamma de baja energía emitida por el pulsar de rayos X Hércules X-1, en lenta rotación. Los fotones de rayos gamma de muy bajas energías formaban un espectro continuo, no uno de líneas, que los investigadores interpretaron como el extremo superior del espectro de rayos X emitido por un disco de acumulación que rodea el pulsar. Sin embargo, aparecía una característica semejante a una línea a una energía de 55 keV (55.000 electronvolt). Había otro segundo rasgo a 110 keV. Ambos llegaron al detector en impulsos a intervalos de 1,24 segundos, que es el período del pulsar. Se cree que constituyen una línea de electrones de ciclotrón junto con un armónico a una frecuencia doble de la del ciclotrón. La existencia del rasgo a 55 keV

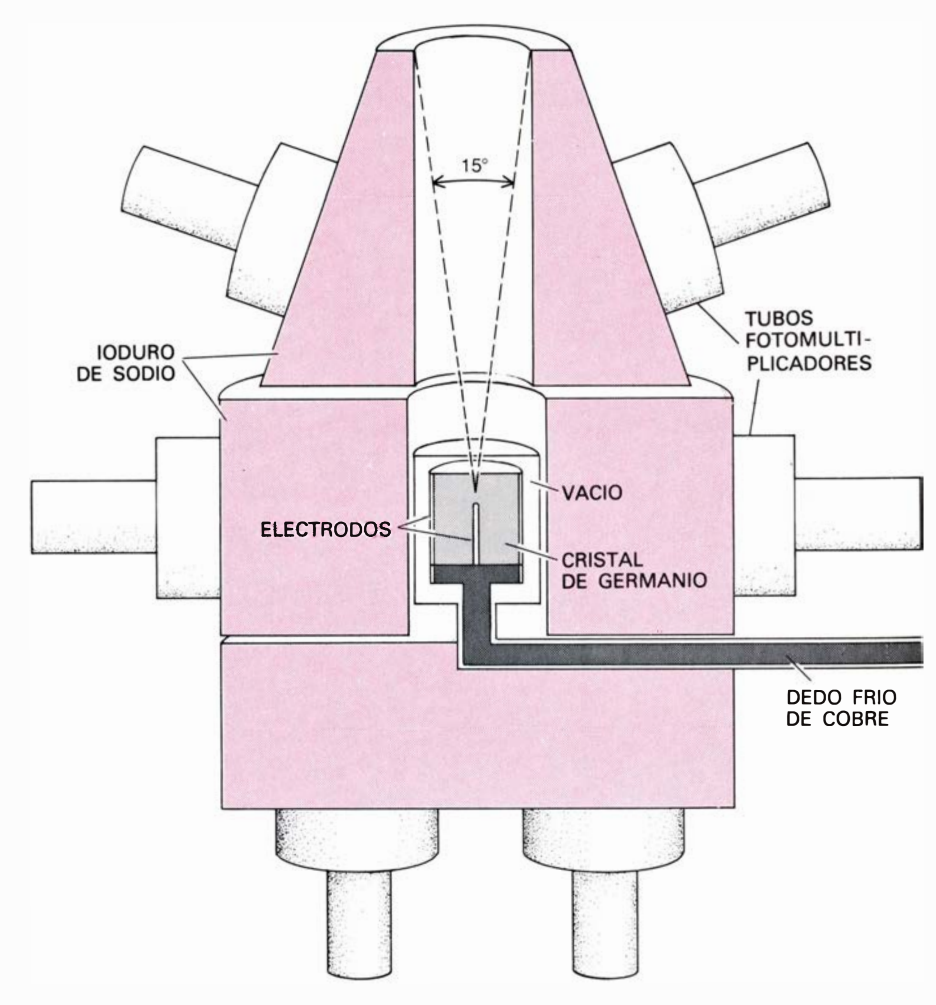
se confirmó en un segundo vuelo con globo y de nuevo el año pasado con un instrumento a bordo del satélite de la NASA *HEAO-1* (el primer Observatorio Astronómico de Alta Energía). Si se trata realmente de una línea de ciclotrón, la observación por el grupo de Trümper hace posible por primera vez medir el campo magnético de una estrella de neutrones. De acuerdo con ello, la intensidad del campo de Hércules X-1 es aproximadamente de $5{,}3 \times 10^{12}$ gauss.

Hay una emisión difusa de rayos gamma que atraviesa nuestra galaxia en todas direcciones y por consiguiente se supone que procede del plano de la Vía Láctea. Su observación requerirá telescopios más sensibles que los existentes en uso hoy día. Se produce porque las partículas de rayos cósmicos (en su mayoría núcleos de alta energía) chocan con los granos de polvo, átomos y moléculas en el espacio interestelar y excitan sus núcleos a estados de energía más alta. El regreso de tales núcleos a su estado normal produce entonces rayos gamma. Como el grano de polvo, átomo o molécula golpeado por una partícula de alta energía retrocede, el núcleo está en movimiento cuando se emite el fotón. El resultado es un corrimiento Doppler en la energía del fotón. El corrimiento tiene lugar hacia un valor más alto si el núcleo se está moviendo hacia el sistema solar, y hacia uno más bajo si el núcleo se está alejando de él. De aquí que para un grupo de núcleos que se están moviendo en forma caótica, las líneas de rayos gamma se ensanchan.

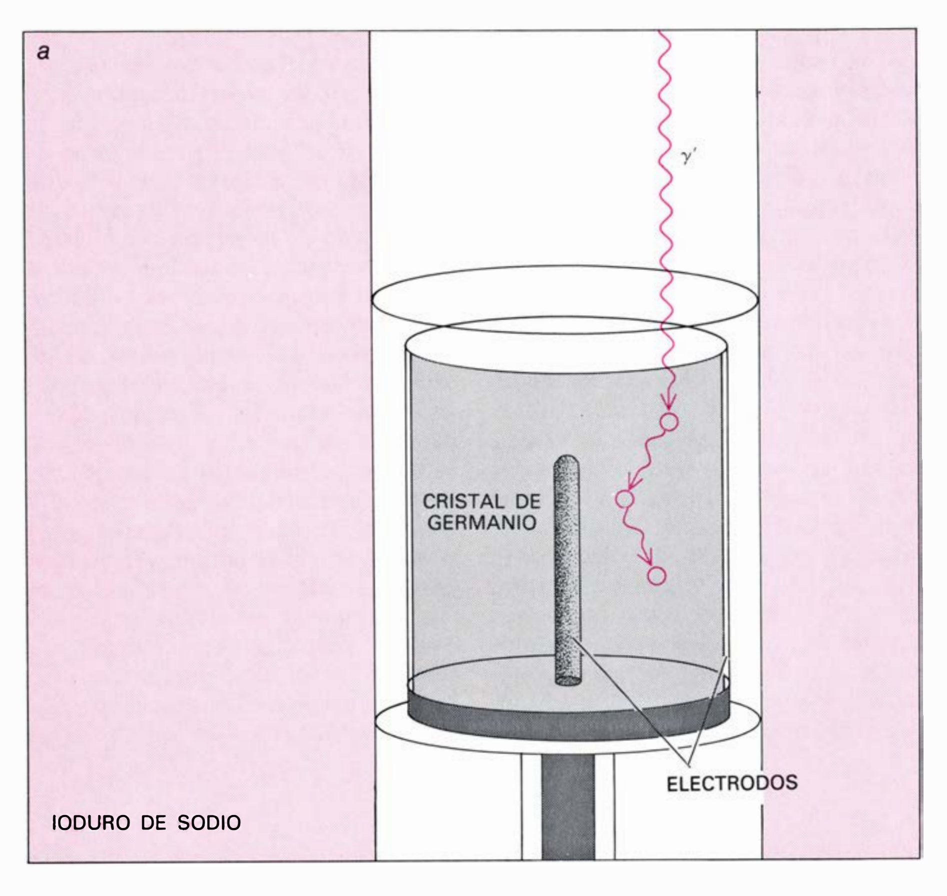
El ensanchamiento es proporcional a las velocidades de los núcleos. Por ello se ha supuesto por Reuven Ramaty, Benzion Kozlovsky, Richard E. Lingenfelter y otros que cada línea de rayos gamma puede ser una combinación de tres contribuciones diferentes. Una componente bien marcada debería originarse de una línea dada cualquiera por efecto de la emisión de fotones por los átomos de los granos de polvo interestelar, que pueden perder su velocidad de retroceso, y después emitir la radiación mientras están todavía dentro del grano. La componente bien marcada de las líneas puede por tanto ofrecer una buena oportunidad para determinar la composición y distribución del polvo interestelar.

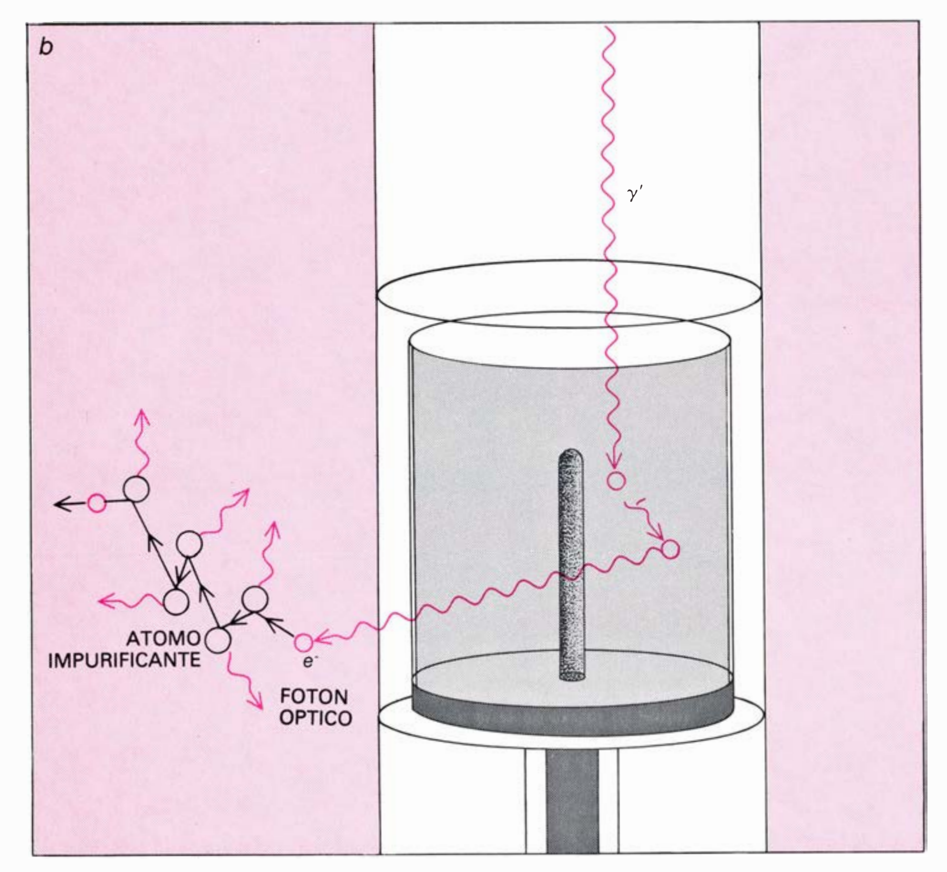
Una contribución algo más ancha para cualquier línea dada puede corresponder a la emisión de fotones por el núcleo de los átomos que no están confinados en granos de polvo. De aquí que la componente más ancha ofrece una buena oportunidad para determinar la composición y distribución del gas interestelar. Otras clases de astronomía pueden servir para el mismo fin, ciertamente. En el caso de la astronomía óptica, sin embargo, el polvo interestelar limita las observaciones hacia el centro de la galaxia a una distancia de solamente 10.000 a 13.000 años-luz en una galaxia cuyo radio es superior a 30.000 años-luz. El medio interestelar es altamente transparente para los rayos gamma. Además, la luz y los fotones de rayos X son emitidos por los electrones en los átomos. Así, el aspecto de las líneas de emisión viene complicado por el estado químico de la materia, tal como su grado de ionización. En cambio, el hecho de que los fotones de rayos gamma sean emitidos desde el núcleo garantiza que su configuración espectral sea independiente del estado químico de la materia.

Finalmente, debe originarse una contribución de espectro ancho sobre cualquier línea dada a causa de la emisión de fotones por los núcleos de los propios rayos cósmicos, cuyas velocidades son una fracción substancial de la velocidad de la luz. La componente ancha, por tanto, ofrece una oportunidad para estudiar los rayos cósmicos, cuyos choques con la materia interestelar se cree son responsables de la existencia de regiones calientes en el gas interestelar. Se sospecha también que los choques de los rayos cósmicos, capaces de fisionar los núcleos de los átomos interestelares, contribuyen a la creación de elementos ligeros tales como el litio y el berilio. (Se pueden postular reacciones de fusión que crean esos elementos en el interior de las estrellas, pero otras reacciones los consumen.) La dificultad para detectar los rayos cósmicos en sí cuando llegan a la Tierra estriba en que dichos rayos



TELESCOPIO DE LINEAS DE RAYOS GAMMA puesto en vuelo por los autores y sus colaboradores; consta de un detector en el interior de otro detector. En el centro está el cristal de germanio, en donde un fotón de rayos gamma incidente es desviado por los electrones. El retroceso de los electrones da lugar a un impulso de corriente eléctrica. El germanio se enfría mediante un "dedo frío" de cobre, enfriado a su vez por nitrógeno líquido; si así no ocurriera, la corriente que fluye espontáneamente enmascararía la corriente generada por un rayo gamma. Rodeando al cristal central hay unos bloques de ioduro de sodio impurificado con una pequeña dosis de átomos de telurio. Aquí, un fotón de rayos gamma incidente es absorbido por un electrón, que choca con átomos del impurificante. La luz emitida por éstos se amplifica por tubos fotomultiplicadores. La detección de un rayo gamma aparece en la página siguiente.





LOS FOTONES DE RAYOS GAMMA se detectan de uno en uno en el telescopio de líneas de rayos gamma de los autores. En *a*, un fotón incidente (α') es desviado repetidas veces, pero permanece en el cristal de germanio central. El movimiento de los electrones que lo desvían da lugar a una corriente eléctrica que representa fielmente la energía del fotón. En *b*, un fotón incidente es desviado hacia el ioduro de sodio circundante, donde produce emisión de luz. El fotón pierde solamente una parte de su energía en el cristal central; por ello las cosas están dispuestas de tal manera que su detección es impedida por una señal resultante del destello de luz en el ioduro de sodio. El suceso *b* es más corriente que el *a*.

interaccionan con los campos magnéticos del sistema solar, de modo que la muestra que llega a la Tierra puede no ser representativa de los rayos cósmicos del espacio interestelar. La detección de su línea característica de rayos gamma eliminaría ese problema.

Otros objetos, asimismo importantes para la astronomía de líneas de rayos gamma, son las galaxias completas o sus regiones centrales. Se cree que los quasars, por ejemplo, son jóvenes galaxias que están inmensamente distantes de la Tierra y son sin embargo notablemente luminosas. Se invocan como explicación dos clases de modelos. Los modelos de la primera clase proponen que en el centro de los quasars se producen explosiones de supernova en gran número. Los modelos de la segunda proponen que las estrellas del centro de los quasars son desgarradas por el campo gravitatorio de un gigantesco agujero negro y que la materia estelar forma alrededor de él un disco de acumulación a alta temperatura. En el segundo caso, es de esperar que puedan observarse las líneas de emisión de rayos gamma procedentes de choques en el disco, con las líneas fuertemente corridas hacia el rojo y además ensanchadas, cada una desde una línea a una ancha y difusa banda espectral, por el campo gravitatorio del agujero negro.

El propio centro de nuestra galaxia es una región misteriosa, debido, en parte, a que los telescopios ópticos no pueden revelar su estructura. En noviembre de 1977, y nuevamente en abril de 1979, elevamos nuestro instrumento mediante globos sobre Alice Springs en Australia, en un esfuerzo por detectar líneas de rayos gamma procedentes de la dirección del centro galáctico. En trabajos previos, un grupo de la Universidad Rice, dirigido por Robert C. Haymes, había encontrado una incierta pero sugerente indicación de la existencia de una línea de emisión cerca de 0,5 MeV. En nuestro vuelo de 1977, detectamos una línea a precisamente 0,511 MeV. Se encontraba hacia la parte central de un fondo continuo de rayos gamma. Confirmamos la existencia de la línea en 1979. Corresponde al aniquilamiento de 10^{43} positrones por segundo en, o cerca del, centro galáctico. La existencia de tal línea en la proximidad del centro galáctico había sido postulada veinte años antes por Philip Morrison, del Instituto de Tecnología de Massachusetts, y más tarde por Floyd W. Stecker, del Centro de Vuelos Espaciales Goddard, y por uno de nosotros (Leventhal).

La literatura científica abunda en sugerencias acerca de la manera cómo se generan los positrones. En lo que quizá sea la más probable vencedora de entre las varias hipótesis, se supone que los positrones son un producto secundario de las supernovas. Se quiere significar con ello que la desintegración de núcleos radiactivos producida en las explosiones estelares libera los positrones, que se alejan del núcleo con energías de cientos de miles de electronvolt. Pueden transcurrir miles de años antes de que un positrón sea aniquilado al encontrarse con un electrón. De aquí que la línea observada desde la Tierra pueda representar explosiones de supernova que se han producido dentro de largos intervalos de tiempo. El hecho de que las estrellas sean muy abundantes en el centro galáctico aboga en favor de la hipótesis.

Una segunda posibilidad es que los rayos cósmicos choquen con la materia interestelar, rompiendo sus núcleos en otros que son radiactivos y emiten positrones cuando se desintegran. Los choques pueden también producir mesones pi, cuya desintegración da lugar a positrones. Sin embargo, cálculos recientes indican que los valores actuales del flujo de rayos cósmicos y de la densidad de materia interestelar son inferiores en dos órdenes de magnitud a los requeridos para explicar la intensidad de la línea de 0,511 MeV que nosotros detectamos. Esta intensidad es grande: es como si la luminosidad del Sol se multiplicase por 10.000 y se concentrase en una frecuencia electromagnética única.

Una tercera posibilidad es que el centro galáctico aloje un elevado número de radio-pulsars que emitan grandes flujos de positrones hacia el espacio interestelar. Otras posibilidades parecen, de momento, menos verosímiles: Una de estas propone la existencia de un gran agujero negro en el centro de la galaxia, con un disco de acumulación en torno a él. Otra postula que en el primitivo universo se crearon gran número de pequeños agujeros negros y que, mediante procesos cuánticos, están, por decirlo así, evaporándose: radian una elevada fracción de su energía en pares electrón-positrón. Quizá la menos plausible de las hipótesis es la que afirma que en el centro de nuestra galaxia haya un gran volumen de antimateria.

Sin embargo, el punto importante es que el mecanismo mediante el cual se generan positrones en el centro de la galaxia se puede determinar por observaciones que podrán realizarse dentro del próximo decenio, e incluso quizá dentro de los próximos cinco años. Se determinará por medio de mapas detallados de la región emisora. Si la región es una fuente puntual, entonces parece que el emisor probable será un agujero negro. Si aparecen máximos próximos a los radio-pulsars, éstos deberán ser los emisores probables. Además, muchas de las hipótesis requieren la existencia de líneas adicionales a la de 0,511 MeV, de modo que la detección de tales líneas limitará las posibilidades.

El punto crucial con respecto a un telescopio de líneas de rayos gamma es que los fotones con energías de rayos gamma no se pueden someter a reflexión con espejos. Penetran la materia y pierden parte de su energía al ser difundidos, o desviados, en su encuentro con electrones. Esta clase de difusión o dispersión rige el proyecto del telescopio: En el centro del telescopio lanzado por nosotros hay un cristal de germanio de un volumen de 130 centímetros cúbicos. Dentro de él, un fotón incidente de rayos gamma típicamente es desviado al chocar con varios electrones en sucesión. El retroceso de los electrones genera una débil corriente cuya intensidad es proporcional a la energía del rayo gamma. Por tanto, los fotones de rayos gamma se pueden detectar uno por uno.

Circundando al detector central hay 180 kilos de ioduro de sodio con átomos de teluro introducidos como impurificante. Este es también un material sensible a los rayos gamma; el choque de un fotón de rayos gamma con un electrón hace que éste retroceda y choque luego con muchos átomos impurificantes, y éstos a su vez emiten impulsos luminosos en forma de fotones con energías ópticas.

La resolución de energía del cristal de germanio a 1 MeV es aproximadamente de 2 keV. Esto significa que los fotones de 1 MeV se pueden detectar como una línea espectral que tiene una anchura de 2 keV. La resolución del detector de ioduro de sodio es aproximadamente 30 veces más baja. Sin embargo, el ioduro de sodio cumple tres importantes funciones. En primer lugar, deja abierto un canal hacia el detector central que define el campo de visión del telescopio. Segundo, protege el detector central de un intenso fondo de fotones de rayos gamma generados en la atmósfera terrestre por rayos cósmicos que llegan del espacio. Tercero, los impulsos luminosos son amplificados por tubos fotomultiplicadores y se usan entonces para impedir la detec-

ción de aquellos fotones de rayos gamma difundidos por el cristal de germanio hacia el ioduro de sodio. Un tal fotón deja algo de su energía en ambos materiales, de modo que la corriente en el detector central no es proporcional a la verdadera energía del rayo.

En el curso del vuelo de un globo que dura uno o dos días, nuestro telescopio recibe órdenes, desde una estación telemétrica de base en tierra, de apuntar en una determinada dirección. Si el telescopio está siguiendo el curso de un emisor de radiación gamma, cada centímetro cuadrado de la superficie del detector central puede recibir un único fotón de rayos gamma de origen astrofísico una vez cada varios minutos. Hay sin embargo más de un cincuenta por cien de probabilidades de que el fotón sea desviado fuera del cristal central. La detección del fotón será por tanto impedida por la casi simultánea producción de luz en la protección circundante. Entonces, lentamente, el detector central acumula un espectro de cuentas de rayos gamma en función de la energía de los fotones. Al final del vuelo, el telescopio se separa del globo a una altura de unos 40 kilómetros y se hace regresar al suelo colgado de un paracaídas.

Nos parece que los experimentos basados en transporte por globo van a continuar siendo importantes en astronomía de rayos gamma. Por un lado, a las alturas a las cuales se lanzan los satélites hay grandes flujos de partículas energéticas presentes en los cinturones de radiación. Las partículas generan un intenso fondo de rayos gamma. Además, los experimentos con globos son menos caros y más fáciles de efectuar que los experimentos con satélites. Desde el momento en que se propone hasta el momento en que empieza a generar datos, un experimento mediante satélite consume de cinco a diez años y cuesta varias decenas de millones de dólares. Un experimento con globo tarda en ponerse a punto menos de dos años y cuesta menos de un millón de dólares. Por otra parte, los satélites están en órbita durante largos períodos de tiempo y acumulan un considerable volumen de datos.

Se ha establecido ahora una red de satélites capaz de determinar precisamente la dirección de los brotes de rayos gamma mediante triangulación. El 5 de marzo de 1979 se detectó el brote de rayos gamma de mayor energía medido hasta ahora por la red, localizado en el centro de los restos de una explosión de supernova en la Gran Nube de Magallanes, una galaxia satélite de la nuestra. P. E. Mazet y sus colegas, del Instituto Físico-Técnico A. F. Ioffe de Leningrado, obtuvieron un espectro del brote. Incluye una característica semejante a una línea a 0,4 MeV: la energía de una línea de aniquilación positrón-electrón tras su corrimiento hacia el rojo. La posición del origen del brote y la detección de la línea sugieren decididamente que lo que está destrás de los brotes de rayos gamma son estrellas de neutrones.

El pasado otoño, la NASA lanzó el satélite *HEAO-3*. Lleva un telescopio de rayos gamma que tiene cuatro detectores de germanio en una envoltura protectora de ioduro de cesio. El instrumento, que fue construido por un grupo en el Laboratorio de Propulsión a Chorro dirigido por Jacobson, está proyectado para funcionar durante un año y explorar la mayor parte del cielo en busca de líneas bien definidas de rayos gamma con una sensibilidad no menor de 10^{-4} fotones por segundo por centímetro cuadrado de superficie de detector. El instrumento está funcionando correctamente, pero es demasiado pronto para tener resultados.

La NASA proyecta ahora lanzar el satélite Observatorio de Rayos Gamma hacia mediados de la década de 1980. El satélite llevará cinco instrumentos sensibles a los rayos gamma, de los cuales tres serán telescopios de banda ancha y, de los dos restantes, uno será un telescopio con base de germanio proyectado por el grupo de Laurence E. Peterson, de la Universidad de California en San Diego, y un telescopio con base de ioduro de sodio proyectado por el grupo de James de Kurfess, del Laboratorio Naval de Investigación; ambos tienen por finalidad la detección de líneas de rayos gamma. En el grupo de Peterson hay investigadores adscritos a los Laboratorios Bell, el Centro de Vuelos Espaciales Goddard, el Laboratorio de Propulsión a Chorro, los Laboratorios Sandia y dos instituciones francesas: el Centro de Investigación Nuclear de Saclay y el Centro de Estudios de Rayos Cósmicos en Toulouse. El grupo de Kurfess cuenta con investigadores de la Universidad Northwestern, la Universidad Rice y del Real Colegio Naval en Gran Bretaña. Los instrumentos deben explorar la mayor parte del cielo durante la vida de dos años que se atribuye al proyecto y alcanzar sensibilidades de línea de 10^{-5} fotones por segundo y por centímetro cuadrado. Con el lanzamiento de tal observatorio, la astronomía de líneas de rayos gamma habrá alcanzado la mayoría de edad.

Ciencia y sociedad

Enlace en Caravaca

Uno de los grandes enigmas de la historia de la Tierra es la extinción, aparentemente simultánea, a finales del Período Cretácico y comienzos de la Era Terciaria, de casi las tres cuartas partes de las especies existentes en aquel momento. Entre los grupos de organismos que desaparecieron se cuentan los dinosaurios, los reptiles marinos gigantes, los reptiles voladores, ciertos invertebrados marinos y virtualmente todo el plancton calcáreo de mares profundos, los organismos cuyos esqueletos de carbonato cálcico formaron los depósitos de caliza. Han sido muchas las hipótesis formuladas para explicar este fenómeno que marca la frontera entre el Período Cretácico y la Era Terciaria, pero hasta ahora ninguna de ellas ha satisfecho adecuadamente las pruebas de que disponemos.

En los últimos años el interés se ha dirigido hacia mecanismos catastróficos, dada la abrupta desaparición de todo rastro fosilizado de restos esqueléticos de determinado plancton calcáreo en el límite Cretácico-Terciaria de las formaciones rocosas de distintas partes del mundo. Entre las hipótesis más aceptadas de las grandes catástrofes se encuentran las que atribuyen la extinción a causas extraterrestres, como la explosión de una supernova cercana o el impacto de un meteorito de grandes proporciones o de un cometa. El interés por estas hipótesis nació, hace algunos años, a raíz del descubrimiento de una concentración extrañamente elevada del elemento traza iridio en un delgado estrato arcilloso identificado con el límite Cretácico-Terciaria en un afloramiento calcáreo próximo a la ciudad de Gubbio, en la Italia central. El exceso de iridio fue descubierto por el físico Luis W. Álvarez, su hijo Walter Álvarez, geólogo, y colaboradores, de la Universidad de California en Berkeley. La medición del elemento traza de las rocas de Gubbio se efectuó en el laboratorio Lawrence en Berkeley, mediante la técnica, extremadamente precisa, del análisis por activación de neutrones.

Al igual que los demás metales del grupo del platino (entre los que se cuentan el osmio y el rodio) el iridio suele escasear en la corteza terrestre. El poco iridio que debió haber en el material acumulado que formó la Tierra se supone que se concentró selectivamente en el núcleo en el transcurso de la diferenciación del planeta. Los meteoritos denominados condritas contienen normalmente 1000 veces más iridio que las rocas corticales, y la composición de las condritas es la que en general se toma como indicativa de la concentración media de iridio en el sistema solar. La presencia de una elevada concentración de iridio en un estrato de roca sedimentaria podría indicar, por tanto, una súbita entrada de material extraterrestre.

Los hallazgos del grupo de Álvarez en Gubbio animaron a otros investigadores a buscar pruebas similares en diferentes lugares. En un reciente número de *Nature,* dos geólogos europeos, J. Smit y J. Hertogen, informaron de los resultados de su estudio de una gruesa secuencia, raramente tan completa, de sedimentos pelágicos calcáreos en Caravaca, al sureste de España. Su análisis revela que la extinción de los fósiles planctónicos en el límite Cretácico-Terciaria "fue abrupta, sin ningún aviso previo en el registro sedimentario, y que el momento de la extinción coincidió con enriquecimientos anómalos de elementos traza, especialmente de iridio y osmio". La escasez de estos dos elementos en la corteza terrestre, añaden, "indica que una fuente extraterrestre, como el impacto de un gran meteorito, pudiera haber proporcionado las cantidades de iridio y osmio requeridas".

En el mismo número de *Nature,* Kenneth J. Hsü, del Instituto Federal de Tecnología de Zürich, adelanta otra interpretación de los datos obtenidos en Gubbio y Caravaca. Sugiere que la colisión de la Tierra con un cometa podría responder al fenómeno transicional del Cretácico a la Era Terciaria. Según Hsü, las pruebas indican que "la extinción... de los grandes animales terrestres estuvo causada por un recalentamiento de la atmósfera durante el impacto del cometa y que la desaparición del plancton marino fue consecuencia del envenenamiento por el cianuro liberado del cometa precipitado y por un aumento catastrófico de la profundidad de compensación de la calcita en los océanos tras la detoxificación del cianuro".

Entretanto, los Álvarez, junto con sus colegas Frank Asaro y Helen V. Michel, han ampliado sus descubrimientos a otros lugares. En un artículo en *Science,* comunican que los resultados de sus análisis por activación de neutrones de calizas de mares profundos expuestas en Italia, Dinamarca y Nueva Zelanda "muestran incrementos de iridio de alrededor de 30, 160 y 20 veces, respectivamente, por encima del nivel de fondo correspondiente precisamente al momento de las extinciones del Cretácico-Terciaria, hace 65 millones de años".

La explicación más probable, tanto de las extinciones como de las mediciones de iridio, según los investigadores de Berkeley, es el impacto de un gran asteroide procedente de la población de asteroides, cuyas órbitas cortan la terrestre. La caída de un objeto de tal magnitud, calculan, habría inyectado a la atmósfera una masa en forma de roca pulverizada equivalente a 60 veces la suya propia. Parte del polvo se habría mantenido durante años en la estratosfera, distribuyéndose por todo el planeta. La oscuridad resultante, escriben, "habría suprimido la fotosíntesis, y las consecuencias biológicas que cabe esperar de ello concuerdan estrechamente con la extinción observada en el registro paleontológico".

Las estimaciones del tamaño del objeto supuesto responsable del fenómeno fronterizo entre el Cretácico y la Era Terciaria son notablemente coherentes. Smit y Hertogen calculan a su hipotético meteorito un diámetro de entre cinco y 15 kilómetros. El cometa de Hsü (que este supone se precipitó en el océano) debió tener 30 kilómetros de diámetro. El grupo de Álvarez ha efectuado cuatro estimaciones independientes del asteroide que proponen; todas ellas dan valores situados en 10±4 kilómetros.

Más veloz que la luz

Durante los últimos diez años los astrónomos han descubierto que cuatro radiofuentes extragalácticas compactas aparentan estar en expansión con velocidades entre dos y veinte veces mayores que la velocidad de la luz. Tres de estas fuentes son quasars, los objetos más distantes y más brillantes del universo, y la cuarta fuente 3C 120, una galaxia. En las longitudes de onda correspondientes a las radioondas cada fuente puede resolverse en al menos dos componentes diferenciadas: una pequeña mancha brillante y un "tubo" difuso cuya longitud triplica o cuatriplica su anchura. Inmersas en ese tubo aparecen otras pequeñas zonas brillantes. Observaciones realizadas a intervalos de meses o de años muestran que tales componentes se desplazan

independientemente. Al combinar las velocidades de separación angular observadas con las estimaciones de la distancia de tales objetos, los resultados parecen indicar que sus velocidades relativas son mayores que la velocidad de la luz.

El principio que afirma que ningún ente físico (como masa, energía o información) puede moverse a velocidad mayor que la de la luz es una consecuencia de la teoría especial de la relatividad. Por supuesto que las cuatro radiofuentes no violan este principio: su velocidad superlumínica debe ser sólo una ilusión. La pregunta que queda por contestar es: ¿qué es exactamente lo que motiva esta ilusión?

Un ejemplo de un fenómeno superlumínico aparente es el efecto de la linterna de flash, en el que la proyección de un haz de luz sobre una superficie distante parece moverse con una velocidad mayor que la de la luz. Por ejemplo, imaginemos una potentísima linterna cuyo haz colimado iluminase una zona circular de una nube distante. Con un movimiento rápido de la muñeca la mancha iluminada se movería a través del cielo a velocidades mayores que la de la luz. Este movimiento no viola la teoría de la relatividad, pues no hay ningún objeto físico que se esté moviendo a velocidad mayor que la de la luz. Los fotones, o partículas de luz, que constituyen la mancha van cambiando de un punto a otro.

Un efecto de linterna constituye la base de uno de los seis modelos desarrollados para explicar esta expansión aparentemente superlumínica de las radiofuentes compactas. Estos modelos han sido analizados por Alan P. Marscher, de la Universidad de California en San Diego, y John S. Scott, de la Universidad de Arizona, en un reciente número de *Publications of the Astronomical Society of the Pacific*.

En el modelo del efecto de linterna se supone que un haz en rotación de fotones u otras partículas emitidas por la radiofuente incide sobre una nube de polvo distante, en la que aparece una mancha brillante debido a las interacciones energéticas de las partículas incidentes con el polvo. Al girar el haz, la mancha brillante puede moverse a través del cielo con una velocidad mayor que la de la luz.

El punto débil del modelo de la linterna radica en la explicación de cómo debería cambiar la velocidad de la mancha brillante con el tiempo. El modelo predice que, excepto para disposiciones altamente improbables del haz y de la nube, la mancha debería decelerar. Tal predicción entra en conflicto con las observaciones de la fuente compacta 3C 345, cuya velocidad ha permanecido constante durante al menos seis años. Un modelo similar, denominado proyecto del eco luminoso, supone un haz de partículas emitidas que incide sobre una serie de nubes con diferentes orientaciones. Con este modelo no se puede dar cuenta de la estructura tubular observada de las radiofuentes.

Otra explicación atribuye la aparente expansión a velocidad superlumínica a una sobreestimación de la distancia de las fuentes. La distancia se deduce del corrimiento hacia el rojo de la radiación recibida, es decir, del incremento de las longitudes de onda debido al movimiento de recesión del emisor. Una hipótesis fundamental en la moderna cosmología afirma que cuanto más deprisa se aleja un objeto, tanto más lejos está. Los corrimientos hacia el rojo de las fuentes superlumínicas son bastante grandes. De acuerdo con la interpretación usual de esos corrimientos, las fuentes están alejándose a velocidades próximas a la de la luz y se encuentran a una enorme distancia del sistema solar.

Algunos teóricos han objetado que los quasars no pueden estar moviéndose tan deprisa como sus corrimientos hacia el rojo implican y, en consecuencia, sugieren que tales corrimientos hacia el rojo no responden a un movimiento de recesión. El que tal indicación sea o no acertada importa poco para la cuestión del mecanismo de las fuentes superlumínicas, puesto que una de ellas no es un quasar sino una galaxia, y para el caso de las galaxias hay bastante seguridad de que el corrimiento hacia el rojo es una medida precisa de su velocidad de alejamiento y de su distancia. Por tanto, la expansión superlumínica de 3C 120 no puede explicarse en función de una sobreestimación de su distancia.

Un modelo que gozó de popularidad inicialmente suponía que en una radiofuente compacta la expansión a velocidad superlumínica podría estar simulada por llamaradas que hicieran erupción aleatoriamente. Desde la Tierra, las llamaradas no aparecerían como sucesos independientes, sino como una fuente de radiación dotada de un rápido movimiento. Este modelo pronto fue rechazado debido a que la expansión observada es demasiado sistemática para poder explicarla en términos de erupciones aleatorias.

Otro modelo sugiere que un cuerpo masivo y difuso, distribuido entre la radiofuente compacta y el sistema solar, podría actuar cual una lente gravitacional, desviando la radiación emitida hasta producir la ilusión de una expansión a velocidad superlumínica. Una tal lente gravitacional constituye hoy la interpretación más verosímil de la peculiar apariencia de otro quasar, donde no hay duda sobre su velocidad de expansión. Sin embargo, para las radiofuentes compactas consideradas, el modelo gravitacional presenta un espinoso problema: 3C 120 se halla lo bastante cerca del sistema solar (500 millones de años-luz) como para que cualquier cuerpo capaz de actuar de lente gravitacional pudiera ser detectado. En ninguna de las fotografías de 3C 120 aparece un cuerpo de esa índole.

El modelo más prometedor para el estudio de las fuentes superlumínicas se basa en la idea de que las partes centrales de las fuentes pueden estar desplazándose a velocidades próximas a la velocidad de la luz, aunque naturalmente sin rebasarla. Martin J. Rees, de la Universidad de Cambridge, ha demostrado que tales velocidades relativistas pueden crear la ilusión de movimiento superlumínico para un observador distante. Dicha ilusión tiene su origen en el fenómeno relativista de la dilatación del tiempo: el tiempo entre dos sucesos acontecidos sobre un sistema de referencia que se está moviendo a velocidad relativista es inferior en el sistema de referencia del propio objeto en movimiento. Dicho de otra forma, si hubiese un reloj sobre la radiofuente compacta, un observador en la Tierra vería el reloj avanzando mucho más deprisa de lo que lo vería avanzar un observador situado en la misma fuente. Como resultado, para un observador terrestre los fenómenos en la fuente parecen tener lugar a un ritmo mucho más rápido que al que ocurren en realidad. Por tanto, la expansión que vista desde la Tierra parece alcanzar una velocidad superlumínica, sería relativista pero con una velocidad inferior todavía a la de la luz cuando se midiera con los relojes, más lentos, de la propia fuente.

Aunque el modelo relativista goce hoy de amplia aceptación, existen varios caminos para modificar las explicaciones restantes que permiten superar sus inconvenientes. Donde todos los astrónomos parecen coincidir es en que la ilusión de movimiento a velocidad mayor que la de la luz es usual en las radiofuentes brillantes y compactas. De diez fuentes estudiadas en detalle, cuatro (y posiblemente también una quinta) presentan expansión a una velocidad aparente mayor que la luz. Los teóricos esperan que aproximadamente

la mitad de todas esas fuentes parezcan expandirse con velocidades mayores que la velocidad de la luz.

Virus cortos

Los virus son parásitos obligados. Sólo sobreviven apoderándose del control de la maquinaria de replicación de una célula viva. Ciertos virus defectuosos ejercen un doble parasitismo. Al haber perdido parte de su genoma en el transcurso de un accidente genético, no se conforman con invadir una célula para reproducirse, sino que precisan además proteínas elaboradas bajo las directrices de un virus normal.

¿Puede un virus "acortado" competir con el virus normal, y con ello reducir la gravedad de una enfermedad? Hace una década, Alice S. Huang, de la Facultad de Medicina de Harvard, y David Baltimore, del Instituto de Tecnología de Massachusetts, lo sugirieron por primera vez. Más tarde, John J. Holland y sus colegas, de la Universidad de California en San Diego, lo ensayaron en ratones.

Nigel Dimmock y sus colegas, de la Universidad de Warwick, informan ahora de un hallazgo aún más positivo. La malignidad del virus del bosque de Semliki, que provoca también una infección fatal en ratones, disminuye en casi el 90 por ciento de los casos en que el animal inhala la modalidad virulenta. (Los virus penetran por los nervios olfatorios.) Una línea de investigación distinta, dirigida por Huang y sus colegas, del Hospital Infantil del Centro Médico de Boston, sugiere una complejidad inesperada en la competencia entre los virus cortos y los largos.

Las modalidades cortas de un virus reciben la denominación de DI, o virus defectuosos interferidores. El nombre recoge sus dos cualidades fundamentales. Su origen en la naturaleza parece que tiene lugar cuando el enzima que dirige la replicación de un virus en una célula del animal hospedante se desengancha prematuramente del virus que está transcribiendo. El fragmento así replicado puede limitarse al punto de fijación del enzima sobre el ácido nucleico del virus. Al carecer de genoma completo, el virus DI resultante no puede reproducirse de la forma habitual. Puede entonces optar por el aprovechamiento de un enzima de replicación elaborado por un virus funcional de extensión normal.

La hipótesis más simple del desarrollo de los hechos dice que la replicación de los virus cortos es más rápida que la replicación de los largos. En consecuencia, en un intervalo dado se elaboran más ejemplares de la modalidad corta, con lo que los virus cortos tienden a monopolizar los enzimas de replicación. De esta forma, los virus cortos pueden impedir la proliferación de los largos, virulentos, y reducir la severidad de una enfermedad.

Dimmock emprendió su investigación "con el ambicioso propósito de determinar de una vez por todas si los virus DI tenían o no una aplicación terapéutica". En un escrito recogido en la publicación inglesa *Spectrum,* apunta la aparición de un "problema desalentador". El ratón debe inhalar un virus DI que pertenezca precisamente a la misma cepa que el virulento. Es más, debe inhalar ambos simultáneamente. Un retraso de sólo unas pocas horas es fatal. La respuesta a si los virus DI son de aplicación médica en enfermedades humanas permanece, pues, en el aire.

Huang y sus colaboradores han determinado mientras tanto la secuencia nucleotídica que codifica la información genética de la modalidad normal y de la modalidad DI del virus de la estomatitis vesicular. De acuerdo con la hipótesis más simple, el virus corto no sería sino una versión recortada del largo. Pero la realidad difiere. El material genético del virus normal es un filamento único de ARN de alrededor de 12.000 nucleótidos de longitud. Salvo tres, los 18 nucleótidos de un final del filamento son complementarios a los 18 del extremo contrario.

Las modalidades DI del virus alcanzan entre 3000 y 6000 nucleótidos de longitud. En parte se corresponden con un fragmento del virus largo. Sin embargo, en las terminaciones del filamento DI parecen haberse corregido los tres nucleótidos que no cuadraban: la secuencia de 18 nucleótidos de una terminación del filamento es justamente el complemento de la secuencia del otro extremo. Es más, la complementariedad parece extenderse más allá de los 18; tiende a abarcar al menos 46 nucleótidos en cada terminal del filamento, y a veces alcanza hasta 60. En algunos virus DI se encuentran numerosas copias de estas secuencias. El éxito de los virus DI pudiera deberse a algo más que a su brevedad y rapidez de replicación. Sus diferencias estructurales podrían facilitar el enlace con el enzima de replicación.

El Arca de Noé

El lugar en que se posó el arca de Noé se encuentra "sobre las montañas de Ararat", sin duda una región de la cabecera del Tigris, dentro de los límites del antiguo Urartu. (Tanto Ararat como Urartu pueden traducirse por "tierras altas".) Tradicionalmente se ha identificado con dos picos gemelos de la Turquía oriental, el Gran Ararat (de 5166 metros de altura, la montaña más alta de la región) y el Pequeño Ararat (de 3926 metros). Desde el comienzo de la era cristiana, si no antes, han mantenido los creyentes la esperanza de hallar los restos del arca en algún punto de estas laderas.

Fernand Navarra, un industrial francés retirado, visitó dos veces el Gran Ararat, en 1955 y en 1969. Recogió una sección de cuaderna trabajada y otros fragmentos de madera en un glaciar de la ladera septentrional de la montaña, a unos 4300 metros de altitud. En una narración del descubrimiento titulada *El arca de Noe: la he tocado* (1974) Navarra afirma que la madera era de roble. Su densidad y color condujo a expertos españoles a declarar que tenía una edad de 5000 años (lo que cuadra bastante con la cronología elaborada por el arzobispo Ussher, quien sitúa la Creación en el año 4004 a. de C.).

Al revisar el tema en un reciente número de *Antiquity,* R. E. Taylor, de la Universidad de California en Riverside, y Rainer Berger, en la Universidad de California en Los Angeles, informan que durante los últimos 15 años varios laboratorios han medido la edad de la madera de Navarra mediante la técnica del carbono 14. Una de las fechas, determinada en el National Physical Laboratory, en Gran Bretaña, corresponde a los años 770-90 d. de C.; la segunda, en la Universidad de California en Riverside, a los años 730-60; la tercera, en la Universidad de California en Los Angeles, al 730; la cuarta, en la Universidad de Pennsylvania, al 640; la quinta, en los Laboratorios Geochron Inc., al intervalo del 620 al 640; y la sexta, también en Geochron, al año 270. Una séptima fecha determinada en Teledyne Isotopes se perdió durante la manipulación de las muestras.

Destacan Taylor y Berger que cinco de las siete determinaciones son estadísticamente idénticas, a la vez que señalan que en los tiempos modernos, junto a la cúspide del Gran Ararat, se han alzado cruces y otros objetos de veneración religiosa. "Es tentador suponer", concluyen, "que uno o dos siglos antes de la conquista islámica del Asia Menor, clérigos armenios o bizantinos erigieran un cenotafio o monumento para conmemorar lo que consideraron... el lugar en que se posó el Arca... Puede que el cenotafio tuviera forma de embarcación."

Monumentos megalíticos

Estos conjuntos de enormes piedras, entre los que se cuenta Stonehenge, abundan por miles en Europa. Su edad fue mucho tiempo incierta, pero ahora han sido fechados en época neolítica

Glyn Daniel

Entre los restos más espectaculares del antiguo panorama cultural de Europa, se encuentran sus muchos monumentos prehistóricos de piedra. Hace tiempo que estos megalitos despertaron el interés y la curiosidad tanto del público general como de los arqueólogos e historiadores de la antigüedad, empeñados éstos en describir correctamente la naturaleza, finalidad, contexto y época de las estructuras en cuestión. Para poner sólo dos ejemplos, los alineamientos de piedras de la región de Carnac, en el sur de la Bretaña francesa, donde más de 3000 menhires se alzan en líneas paralelas que se extienden en un trecho de unos seis kilómetros, y el "Grand Menhir Brisé" de la cercana Locmariaquer, roto hoy en día, pero originariamente de 22 metros de largo, figuran entre las reliquias más notables de la Francia prehistórica.

Sin duda el más famoso de todos los monumentos megalíticos es Stonehenge, en la llanura de Wiltshire, al sur de Inglaterra. Visitado por miles de personas todos los años, sólo lo supera la Torre de Londres como atracción turística. Cuenta con más bibliografía que ningún otro yacimiento arqueológico del mundo, incluidas las Pirámides de Egipto y las grandes estatuas de la isla de Pascua, y más también que yacimientos míticos como la Atlántida. La cantidad de libros acerca de Stonehenge y de otros monumentos megalíticos que han salido de las prensas en el pasado decenio, aproximadamente, dan la medida del continuo interés que despiertan estas antigüedades.

Pero es también, ¡ay!, una clarísima prueba de la imaginación, del desvarío y la credulidad de muchos autores y de la abismal ignorancia de muchos pretendidos arqueólogos que sólo pueden ser calificados, aunque ello no sea muy caritativo, como fantasiosos e ilusos. El fenómeno no es nuevo. En fecha tan remota como 1911, el libro de G. Elliot Smith, *The Ancient Egyptians,* traía todos estos monumentos europeos antiguos de las riberas del Nilo. Tales ejercicios de imaginación continúan. En año tan reciente como 1977, Euan Mac-Kie, en *The Megalith Builders,* declaraba que eran obra de sabios procedentes del Egipto predinástico y de Sumer. Hay otros, entre ellos Erich von Däniken, que contemplan a los constructores megalíticos como viajeros del espacio. También está ahora muy difundida la creencia de que estos monumentos fueron construidos con un propósito astronómico, y palabras tales como "astroarqueología" y "arqueoastronomía" se pronuncian con plena libertad. Echemos, sin embargo, una mirada rigurosa y equilibrada a estos edificios en función del conocimiento detallado que tenemos de la Europa antigua.

El estudio de los megalitos

En 1849, en un libro titulado *Cyclops Christianus,* un profesor de Oxford, Algernon Helbert, acuñaba la palabra megalito (del griego *megas,* grande, y *lithos,* piedra). La palabra hizo fortuna. Aunque en 1872 James Fergusson, un escocés experto en arquitectura, puso por título a un libro suyo *Rude Stone Monuments in All Countries* ("Monumentos bastos de piedra en todos los países"), también él hablaba de megalitos. Y lo mismo T. E. Peet en sus *Rough Stone Monuments* ("Monumentos toscos de piedra"), publicado 40 años después.

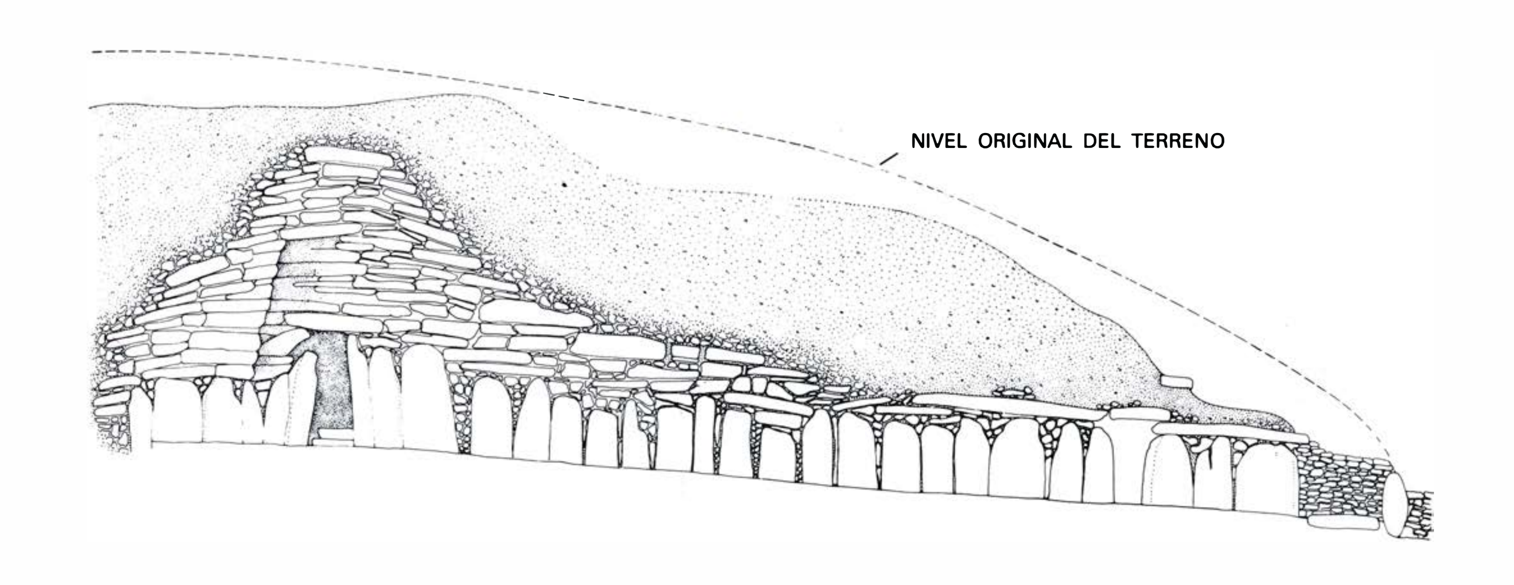
Estas obras iniciales pusieron los cimientos del verdadero estudio de los megalitos, pero introdujeron un error fundamental. A medida que viajeros como Fergusson recorrían países no europeos, encontraban grandes monumentos de piedra en Argelia, Palestina, Etiopía y Sudán, el Cáucaso, Persia, Beluchistán, Cachemira y la India central y meridional. (En la India los monumentos megalíticos del Decan, muchos de ellos con "ojos de buey" parecidos a los de las tumbas megalíticas de Europa, interesaron al principio a Fergusson a la hora de emprender sus estudios comparativos.) La lista no acaba aquí. Hallamos megalitos en Assam, en Sumatra y en algunas islas del pacífico como Malekula, en las Nuevas Hébridas. (Las figuras de piedra de la isla de Pascua, aunque sean desde luego grandes, no son megalitos en el sentido generalmente admitido.) En Japón se construyeron tumbas megalíticas desde el siglo II a.C. hasta el VII d.C., en que el emperador Kotoku las prohibió como trabajo baldío. Las civilizaciones precolombinas del Nuevo Mundo también practicaron la construcción megalítica.

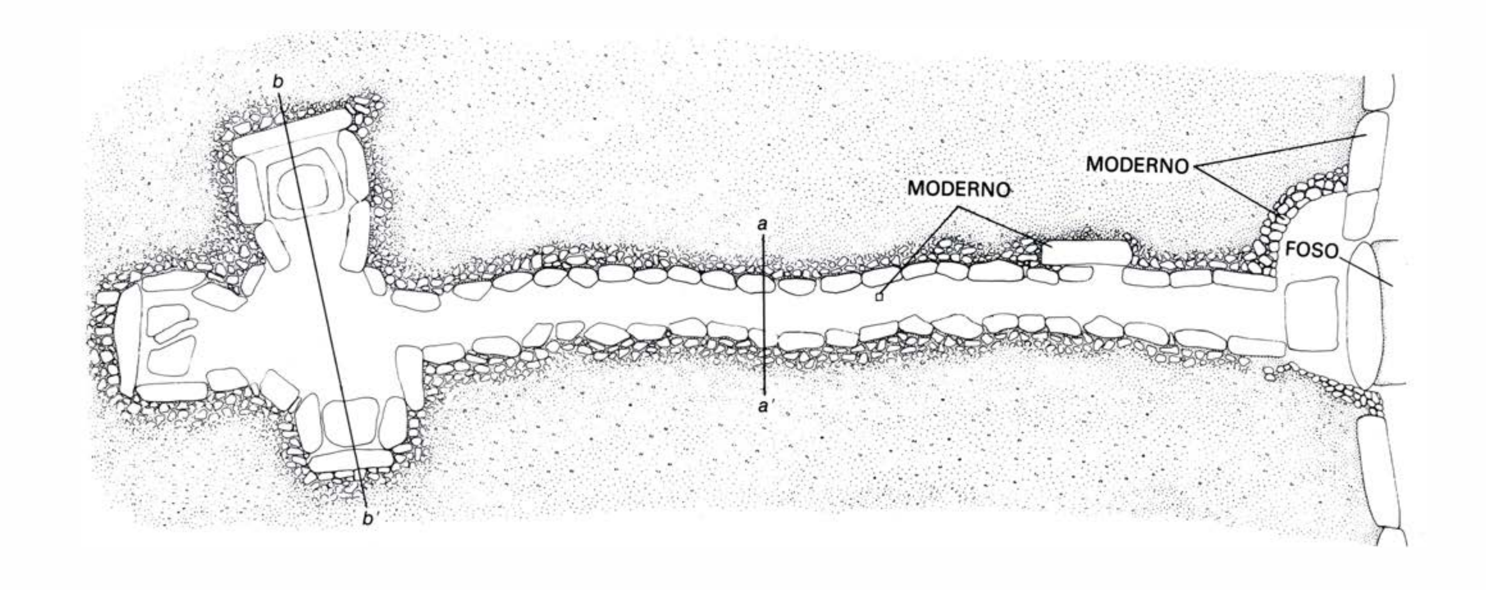
El resultado de aplicar el término megalítico a todos estos monumentos de diferentes países, de diferentes épocas y con diferentes contextos culturales dio lugar a un absurdo: a la idea de que los edificios guardaban una conexión genética, que eran obra de una raza megalítica o de un pueblo megalítico. Hace tiempo que este concepto ha sido abandonado y hoy todo el mundo sabe que los edificios megalíticos de las diferentes partes del mundo son parecidos porque están hechos de materiales parecidos y de modos parecidos. Los paralelismos son particularmente llamativos en las cámaras megalíticas o habitaciones que encarnan los principios básicos de lo que se conoce como arquitectura arquitrabada (del latín *trabes,* viga). Este tipo de edificio es como levantar una casa de naipes o de tacos de niño: se hincan en el suelo losas de piedra (ortostatos) y se ponen otras encima de ellas como dinteles. Las posibilidades arquitectónicas son limitadas, y así no sorprende que una cámara megalítica de Francia o de Irlanda, y que data del tercer milenio a.C., se parezca a una cámara megalítica del sur de la India fechable a fines del primer milenio a.C.

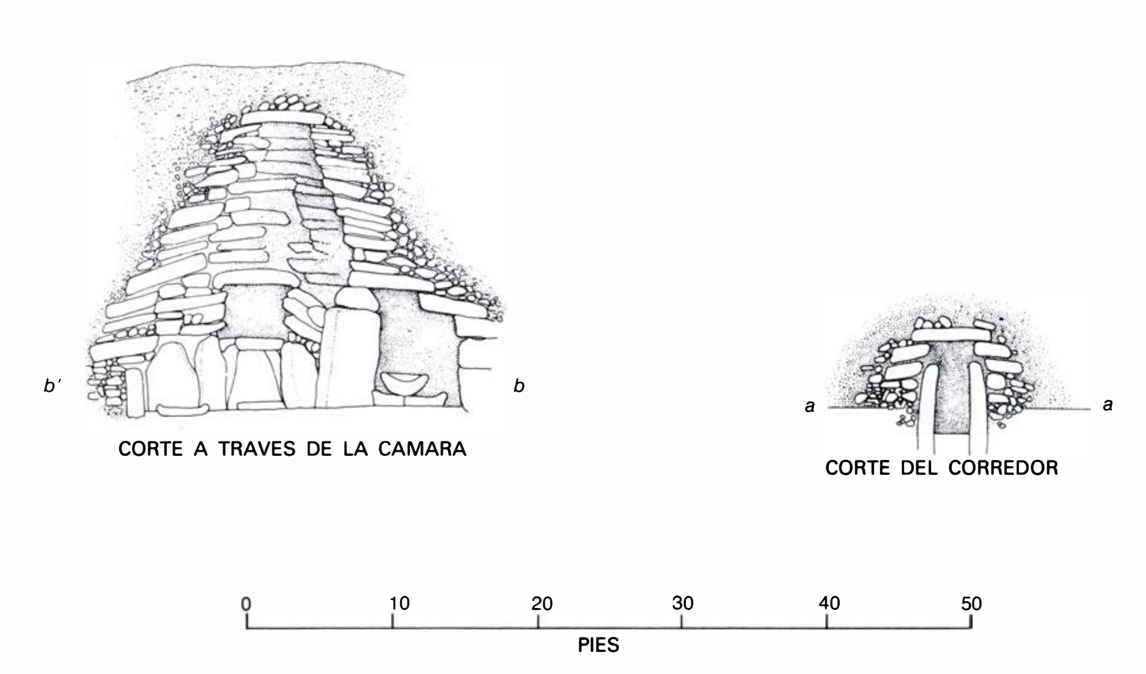
Otra confusión muy común es la existente entre arquitectura megalítica

GRAN FORMACION DE CARNAC en la Bretaña meridional. Comprende tres avenidas separadas de menhires. En esta fotografía aérea se ve el despliegue de Kermario, consistente en 10 hileras de menhires, tallados en el granito local y colocados en una extensión de más de 1000 metros.

EL MAS FAMOSO MONUMENTO del legado megalítico, Stonehenge, es un anillo rodeado de un terraplén y de un foso, situado en el llano de Wiltshire, al sur de Inglaterra. El anillo fue construido por etapas, a partir del 2800 antes de Cristo, aproximadamente, y terminado algo después de 1100 a.C.







LA MAYOR CAMARA FUNERARIA del legado megalítico es Newgrange, en el Condado de Meath, en Irlanda, que aparece en alzado en la parte superior y en un plano y secciones en la inferior. El estrecho corredor que conduce a la cámara de enterramiento del centro del gran túmulo mide unos 23 metros de longitud, y la bóveda hecha por aproximación de hiladas de la cámara alcanza unos siete metros de altura. Se construyó hacia 2500 a.C.

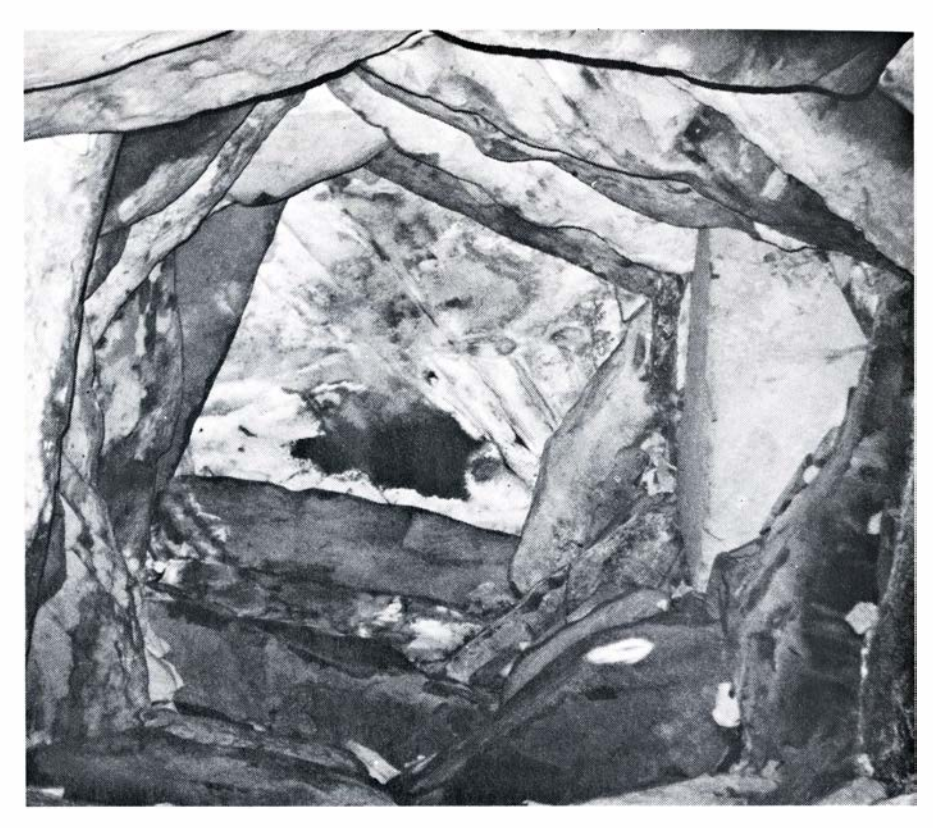
y arquitectura ciclópea. Esta última emplea también piedras grandes, pero un edificio ciclópeo está construido de piedras que han sido cuidadosamente ajustadas, aunque sean de forma irregular, y generalmente puestas en hiladas. La arquitectura ciclópea se encuentra tanto en el Nuevo Mundo (por ejemplo, los edificios de los incas en el Perú) como en el Viejo (por ejemplo, las ciudadelas de Micenas y de Tirinto, en Grecia, o los nuraghi, o torres de piedra, de Cerdeña).

Los megalitos europeos

Los edificios megalíticos de Europa se dividen en cuatro categorías fundamentales. La primera es la del menhir, o piedra aislada erecta. La palabra viene del galés *maen,* piedra, e *hir,* largo. Bretaña es rica en menhires cuya longitud oscila de uno a seis metros. Una notable excepción la constituye el "Grand Menhir Brisé", una piedra acostada mucho más grande. Nadie sabe si esta enorme piedra estuvo en pie alguna vez; las descripciones más antiguas hablan de ella tal como está ahora.

Una clase especial de menhir, conocida como estatua-menhir, está esculpida de manera que muestra la efigie de una persona, humana o divina. Estas se encuentran en el sur de Francia y en el norte de Italia, con ejemplos periféricos tanto en España como en las islas del Canal. Estos interesantes pies derechos, la más primitiva escultura de bulto a escala monumental de la historia humana, pueden datarse a fines del tercer milenio a.C. y comienzos del segundo. [El autor pasa por alto, sin duda distraídamente, las muchas estatuas egipcias y mesopotámicas anteriores a las fechas que él asigna a las estatuas-menhires. Para poner un ejemplo que no admite discusión: la Esfinge de Gizeh, muchas veces reconocida como la escultura más célebre del mundo, data de los tiempos del faraón Kefrén, que reinó entre 2521 y 2495 a.C., según la llamada hoy cronología corta. N. del T.] No están, sin embargo, necesariamente relacionados con los menhires no decorados.

La segunda categoría de construcciones megalíticas se compone de piedras erectas agrupadas. Las piedras están dispuestas bien en filas, como lo están las de Carnac, o en lo que se solía llamar círculos de piedras. Los cuidadosos levantamientos de Alexander Thom, profesor retirado de ingeniería de la Universidad de Oxford, han demostrado que muchos de estos últimos monumentos no son estrictamente circulares; muchos tienen la forma de una elipse o de una elipse aplanada. Como consecuencia, el término que ahora se está poniendo en boga para designar estos recintos megalíticos es el de anillo de piedra. Algunos están aislados; otros rodean un túmulo funerario. Algunos están asociados con filas de piedras; otros, como los famosos monumentos de Stonehenge y Avebury, en el sur de Inglaterra, y Brodgar, en Orkney, están rodeados de fosos y terraplenes.

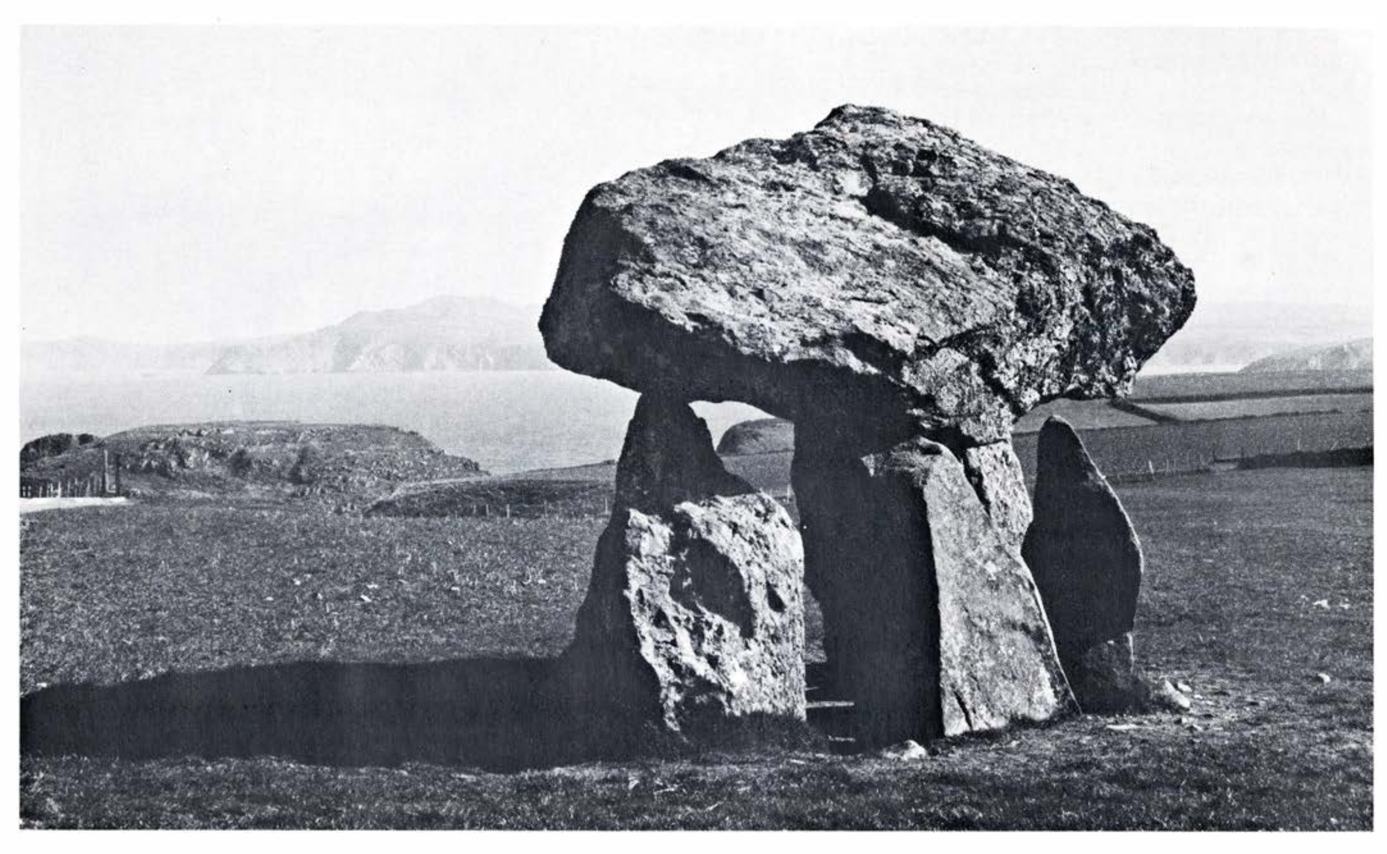


MIRANDO A LO ALTO DE LA BOVEDA FALSA de Newgrange, esta fotografía revela su construcción escalonada. Las sucesivas capas de losas se fueron colocando cada vez más próximas hasta que las piedras de coronamiento pudieron salvar el hueco que se había ido estrechando (*véase la ilustración detallada de la página opuesta*). Una bóveda falsa similar a esta misma fue construida en Maes Howe, en Orkney.

La tercera categoría, la cámara de enterramiento, constituye la forma más generalizada de monumento de piedra europeo. Deben quedar restos de unas 40.000 a 50.000 de estas cámaras, y originariamente quizás hubiese el doble. La mayor parte de ellas se encuentran en España y Portugal, Francia, Gran Bretaña e Irlanda, Escandinavia y norte de Alemania. Unas están completamente enterradas bajo grandes túmulos de tierra y piedra, otras conservan señales de túmulos destruidos en parte y, por último, las hay completamente exentas. En el siglo XIX se creía que todas las cámaras exentas eran restos de cámaras en túmulos, pero ahora no se piensa así. Algunas son efectivamente los restos de túmulos sepulcrales desmantelados, pero otros, como las cámaras de portal de Irlanda y las grandes galerías de piedra del oeste de Francia, estuvieron probablemente siempre tal y como hoy están.

Muchas cámaras funerarias despliegan una técnica de cubrición más compleja que la simple losa de coronamiento que en adelante llamaremos también dintel; sobre los ortostatos verticales se colocaban piedras, de tal modo que se superpusiesen unas a otros hasta formar un techo acampanado, o bóveda. Entre las más famosas bóvedas hechas por aproximación de hiladas que hoy sobreviven se encuentran Maes Howe, en Orkney, y Newgrange, en el Condado de Meath, en Irlanda. Maes Howe fue violada por los vikingos, pero Newgrange, con su bóveda levantada a siete metros del suelo se ha mantenido incólume durante 4500 años. Es una de las maravillas de la Europa prehistórica.

Muchas de las piedras de estas cámaras son muy grandes. Se calcula que la piedra de coronamiento de la cámara de Mount Browne, en el Condado de Carlow, en Irlanda, pesa unas 100 toneladas. La gran cámara megalítica de Bagneux, cerca de Saumur, en el oeste de Francia, tiene unos 20 metros de largo por cinco de ancho, y su techo de cuatro dinteles tiene tres metros de altura, espacio suficiente para cualquiera que entre en ella. (Hubo un tiempo en que albergó un café.) Los dinteles miden unos 60 centímetros de



DOLMEN O "MESA DE PIEDRA" es el nombre dado en Bretaña a este tipo de monumento megalítico. Se compone de una sola piedra de cubrición apoyada en tres o más soportes verticales, y es la forma más sencilla de cámara funeraria. Este monumento, situado en Pembrokeshire, en Gales, es conocido como el cromlech de Carreg Samson. (En galés cromlech significa dolmen.) Algunas cámaras están al descubierto; otras, bajo un túmulo.

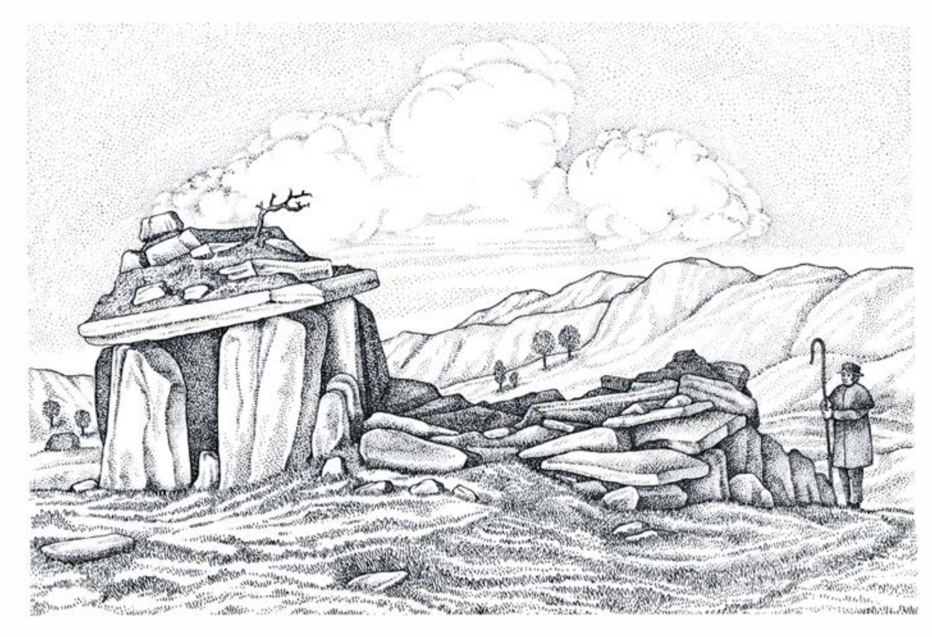
CAMARA MAYOR, también en Pembrokeshire, que ahora sólo tiene una piedra de cubrición. Esta cámara funeraria exenta, situada en Pentre-Ifan, fue catalogada por George Owen en su exploración de Pembrokeshire, publicada en 1603. En Europa debe haber unas 50.000 de éstas.

espesor; se calcula que el mayor de ellos pesa 86 toneladas.

La cuarta categoría, el templo megalítico, es más reducida. La mayoría de sus ejemplos se localizan en el Mediterráneo, en la isla de Malta y en su vecina Gozo. La aplicación del término "templo" a estos grandes monumentos megalíticos puede evocar en la mente de algunos un edificio de aspecto comparable a las construcciones columnadas del Egipto dinástico o de la Grecia clásica. Los monumentos malteses son muy distintos. Sus paredes sólidas se componen de losas muy grandes de piedra; sus plantas incluyen ábsides salientes y probablemente estuvieron techados con vigas de madera y paja. Desde luego se encuentran entre los más impresionantes monumentos arquitectónicos prehistóricos de toda la zona mediterránea.

No puede caber duda de que los monumentos malteses fueran templos. No contienen ni tumbas ni restos de ocupación doméstica; pero sí, por el contrario, muchos objetos de culto, incluso estatuillas de una divinidad femenina y piedras decoradas con espirales y otros motivos. Tampoco hay duda de que datan del cuarto milenio a.C. Son, por tanto, el ejemplo más primitivo de arquitectura en piedra en el Mediterráneo, y junto con las cámaras megalíticas de España y Francia, del mundo entero.

Durante el Medievo el pueblo llano, y los hombres cultos también, creían que estos grandes monumentos de piedra eran obra de gigantes de antaño. A decir verdad, la primera excavación de una tumba megalítica de que hay constancia en Suecia tuvo el objeto específico de poner a prueba esta suposición. Se comprobó que era errónea: los restos hallados en la cámara no eran los de gigantes. Historiadores de la antigüedad más recientes trataron de explicar los megalitos en función de la historia escrita. Los edificios fueron atribuidos a varios presuntos constructores: los romanos, los anglosajones, los daneses, los godos y los hunos. Herbert, por ejemplo, el acuñador de la palabra megalito, aducía que como Stonehenge y otros megalitos británicos no eran mencionados por historiadores ni visitantes romanos de Britannia, los edificios debían ser posteriores a los romanos.



LA EROSION O EL VANDALISMO habían desmantelado el túmulo que en su tiempo ocultaba esta tumba de cámara en la época en que el artista diseñó este monumento en la década de 1840. Sólo una de las piedras de cubrición sostiene aún algo de la antigua cubierta de tierra y cascajo. El grabado en que se basa este bosquejo de Bryn Celli Ddu, de Anglesey en Gales, apareció en *Archaeologia Cambrensis* (1847).

El quién y el por qué

Paulatinamente se llegó a aceptar que los megalitos eran prerromanos. Los habitantes prerromanos del noroeste de Europa, esto es, los antiguos galos y britanos, habían sido descritos, por César y por otros, como un pueblo que tenía una clase de sabios llamados druidas. No ha de sorprender, por tanto, que los megalitos llegaran a ser atribuidos a los druidas. Algunos interpretaron incluso las cámaras exentas como altares de los druidas.

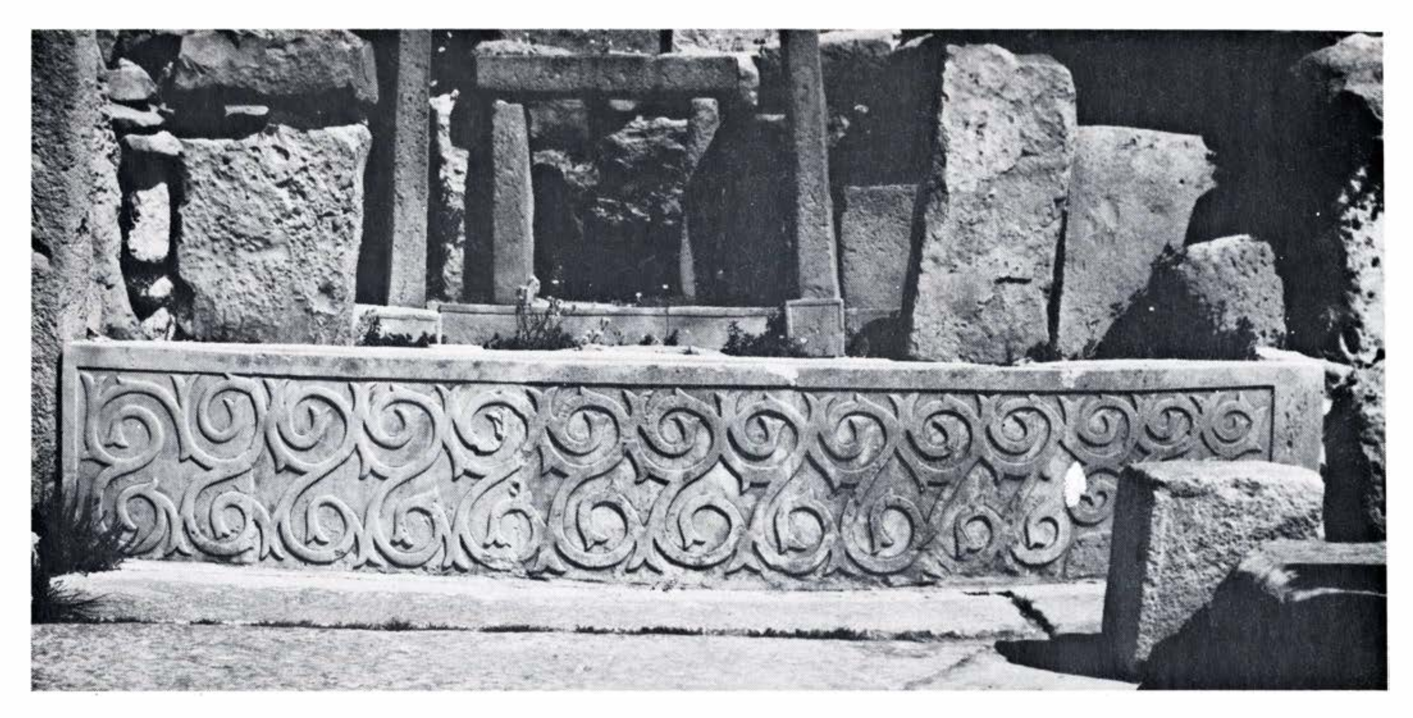
A lo largo de los dos últimos siglos, el crecimiento de las excavaciones arqueológicas ha revelado que las cámaras megalíticas eran en primer lugar sitios de enterramiento, usados colectivamente por una comunidad o por una familia durante cierto período de tiempo. Algunas grandes cámaras han proporcionado los restos de 200 individuos o más. Sin embargo, no todas las cámaras megalíticas eran necesariamente tumbas. Algunas, como las cámaras exentas de Irlanda y del valle del Loira en Francia, pudieron haber sido templos, aunque difícilmente templos druídicos. A decir verdad, la línea de separación entre un templo y una tumba no es estricta ni infranqueable, como se comprueba examinando las iglesias y catedrales cristianas.

Los ajuares funerarios encontrados en muchas tumbas megalíticas son poco llamativos: cacharros sin decorar y hachas de piedra pulimentada. En algunas regiones, sin embargo, son ricos y notables. En Escandinavia se encuentran vasijas decoradas con *oculi*, u ornamentos en forma de ojos. En España muchas tumbas proporcionan elevadas cantidades de placas de pizarra cubiertas de adornos, algunos de ellos antropomorfos. Las tumbas de la Bretaña francesa han proporcionado hachas pulimentadas exquisitamente fabricadas y anillos de jade, granate y otros materiales preciosos. Estos últimos datan del tercer milenio a.C., adelantándose mucho a lo que en jade hicieron los chinos y los mayas.

La finalidad de los menhires es más difícil de precisar. A veces se encuentran enterramientos a sus pies; pero no son tumbas. Además, esos enterramientos ni siquiera pueden proporcionar una fecha al monumento, y mucho menos señalar su finalidad original. Es posible que estas piedras erguidas tuviesen el destino de lindes territoriales. También lo es, por el contrario, que sirviesen para recordar a los muertos, como una versión prehistórica del cenotafio, o tumba vacía, de los griegos.

El objeto de los alineamientos de menhires y de los anillos de piedra no es evidentemente ni funerario ni doméstico. Quizás estos monumentos fueran lugares de asamblea donde tribus o comunidades se reunían de vez en cuando, tanto con fines seculares como religiosos. Yo me los imagino funcionando como los *pardons* bretones de la actualidad: ocasiones en que los sacerdotes celebran ritos sagrados, en que son bendecidas las cosechas y las bestias, en que amigos y parientes se reúnen y disfrutan de una ceremonia social, un mercado y una feria, especialmente una feria de contratación.

¿Se ajusta esta descripción a Stonehenge, el anillo de piedras más conocido entre los de su especie? En primer término, habría que recordar que Stonehenge es un complicado monumento



ESPIRALES CONTRAPUESTAS decoran una pared baja en un templo megalítico de Malta en Hal Tarxien, cerca de Valetta. Los numerosos templos megalíticos de Malta y de la cercana Gozo, considerados antes de inspiración minoica, han revelado ahora poseer un origen autóctono.

LA DECORACION DE ESPIRALES predomina entre los motivos diseñados en bajorrelieve en esta columna megalítica ubicada a la entrada de New Grange. La decoración, si es tan antigua como la cámara misma, es por lo menos 1500 años posterior a los templos megalíticos malteses más antiguos.

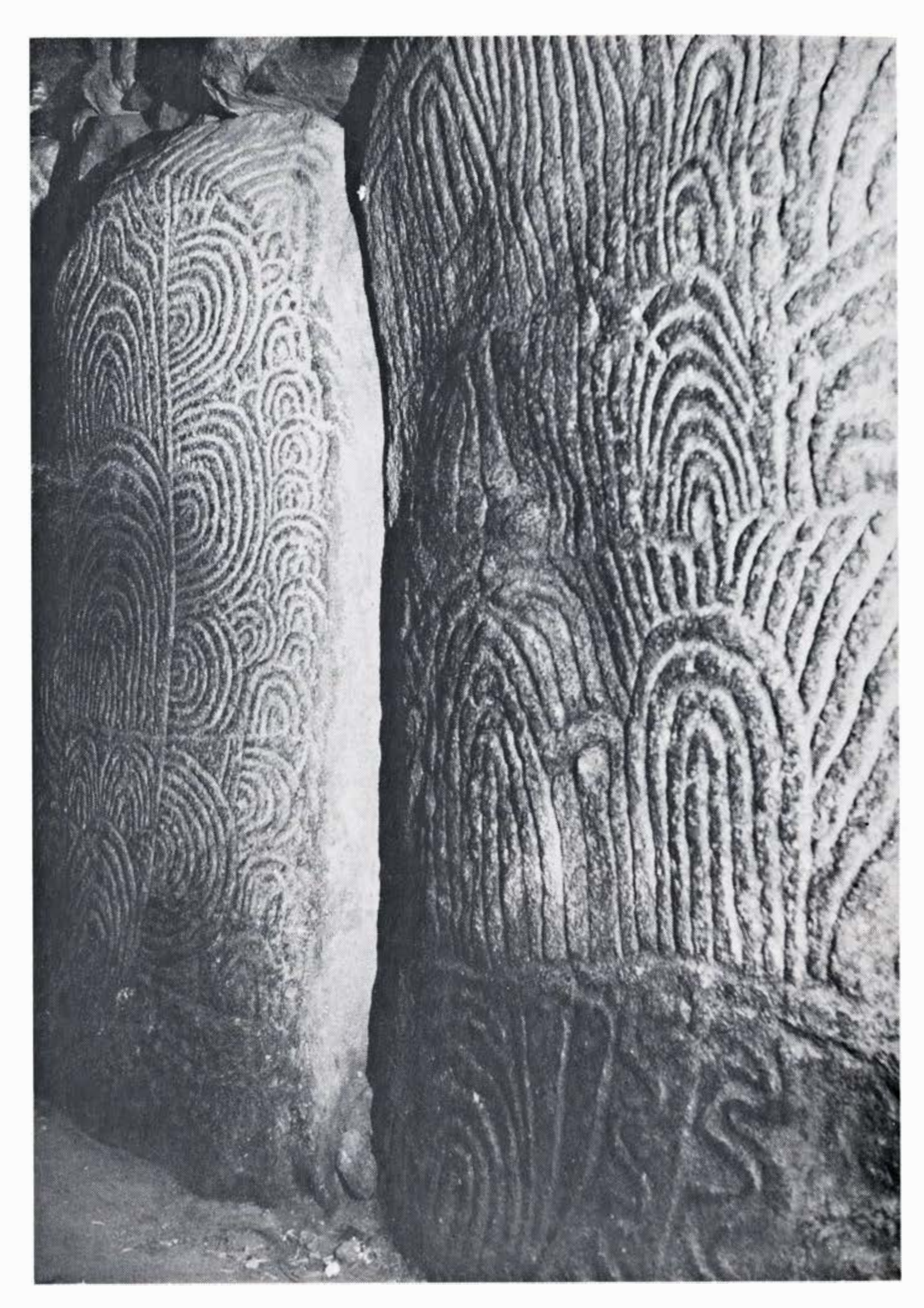
de varias épocas y de muchos elementos arquitectónicos, incluidos dos círculos, dos herraduras y "trilitos" cuidadosamente conformados, que están a gran distancia de los simples ortostatos y dinteles. Las piedras mismas son en parte de arenisca de los "Downs" de Marlborough y en parte piedra azul "forastera", de las colinas de Preseli, de Pembrokeshire, en el sur de Gales. Hay unas 80 de estas piedras azules, cada una de ellas de cuatro toneladas de peso; la distancia de Preseli a Stonehenge es de 193 kilómetros a vuelo de pájaro. Las piezas de arenisca, más conocidas como *sarsen stones,* son también grandes y están cuidadosamente labradas. El componente individual de mayor tamaño de los trilitos de *sarsen* mide cerca de 10 metros, y la altura media de las piedras del círculo de *sarsen* es de cuatro metros. Uno de los recursos más interesantes y refinados, desde el punto de vista arquitectónico, empleados por los constructores de Stonehenge es lo que en arquitectura griega se conoce como éntasis. Este consiste en dar forma a un pie derecho de modo que quede anulado el efecto de la perspectiva; cuando las piedras se miran de cerca no parecen disminuir hacia arriba sino que producen la ilusión óptica de tener los lados paralelos.

La medición del año

No cabe ninguna duda de que Stonehenge fue construido con su eje central apuntando a la salida del sol en el solsticio de verano. Hace muchos años el astrónomo británico Sir Norman Lockyer demostró que calculando la exacta orientación de Stonehenge podía atribuir una fecha precisa a la construcción del monumento. Su conclusión fue 1680 a.C. (± 200 años). De entonces acá, R. J. C. Atkinson, del University College de Cardiff, que ha hecho más que nadie con la excavación y la investigación para contribuir a nuestro moderno conocimiento de Stonehenge, ha señalado que, debido a errores del razonamiento original de Lockyer, esta fecha debiera modificarse ligeramente. La nueva lectura de los datos arroja el 1840 a.C. (± 200 años).

La finalidad astronómica de Stonehenge y de otros anillos de piedra no fue discutida en serio hasta después de la Segunda Guerra Mundial, cuando Gerald S. Hawkins, de la Universidad de Boston, sostuvo en su libro *Stonehenge Decoded* ("Stonehenge descifrado") (1966), que el monumento era una calculadora gigante para la predicción de eclipses, tanto lunares como solares. Cinco años más tarde, Thom, en su libro *Megalithic Lunar Observatories,* postulaba que muchos monumentos megalíticos fueron utilizados para la observación de los movimientos y fases de la luna; Thom había llegado a sostener, en sus postulados, la existencia de una "yarda" megalítica, que medía 2,72 pies, y a suponer que los constructores de los anillos de piedra tenían conocimiento de la geometría pitagórica 2000 años o más antes de los griegos. Todas estas son pretensiones extravagantes y nada convincentes; lo único que tenían los constructores de megalitos era un conocimiento práctico del trazado de triángulos rectángulos.

Mucha gente, aburrida sin duda de la prosaica interpretación de los megalitos que se saca de la investigación arqueológica, se ha encaramado al carro de la farsa de Hawkins-Thom, aceptando a los constructores de megalitos no sólo como expertos en geometría pitagórica y poseedores de exactas unidades de medición, sino también como hábiles astrónomos que estudiaban los eclipses, los movimientos de la luna y las posiciones de la estrellas. Para mi esto es una especie de versión académica refinada de la arqueología astronáutica. Los paladines de la arqueoastronomía, aunque muy discretamente excluyen a los sabios del espacio exterior, insisten sin medida en la presencia en la Europa antigua de sabios con una pasión aparentemente



MC VO SEMICIRCULAR predominante en la labra en bajorrelieve de estos pies derechos de la cám a funeraria megalítica de Gavrinis, en Francia, aproximadamente contemporánea de la irlandesa Newgrange. No se sabe si estos motivos eran decorativos o quizá tenían un claro significado religioso.

religiosa por la astronomía. En mi opinión, la tesis de la interpretación de los monumentos megalíticos como observatorios astronómicos no ha sido nunca demostrada. Las interpretaciones parecen subjetivas e impuestas por el observador. Nuevos levantamientos planimétricos están demostrando ya la inexactitud de algunas de las observaciones anteriores y socavando las esperanzas de aquellos que creen que los constructores de megalitos eran esclavos de un culto astronómico.

Todo el estudio de los megalitos de Europa ha experimentado una revolución, no provocada por quienes lo hacen sin quitar ojo de la luna y de las estrellas, sino por los adelantos en la datación prehistórica. El primero de éstos fue la invención, por Willard F. Libby y sus colegas, de la datación por el carbono 14. Las primeras fechas del carbono 14 fueron publicadas hace 30 años, y desde entonces miles de ellas han sido fijadas por los laboratorios de todo el mundo. Una técnica geocronológica más reciente, la datación por termoluminiscencia, ha confirmado muchas determinaciones del isótopo carbono 14.

La cuestión del cuándo

Antes de la revolución del carbono 14 (de cuya incidencia revolucionaria en el campo de la arqueología prehistórica no cabe dudar), las edades de varias obras antiguas del hombre habían sido adivinadas o calculadas, de un modo penosamente incierto, por medio de correlaciones entre las civilizaciones fechadas de Egipto y de Mesopotamia y la Europa bárbara sin fechar. Así, en el período comprendido entre las dos guerras mundiales, era habitual atribuir los megalitos europeos más antiguos al período neolítico, esto es, entre 2500 y 2000 a.C., y los grandes monumentos tales como Newgrange y Stonehenge a la Edad del Bronce, entre 1800 y 1500 a.C.

Ahora, gracias a la revolución del carbono 14, podemos afirmar confiadamente cuál es la verdadera antigüedad de los megalitos de Europa. Los templos de Malta datan del 4000 al 2000 a.C. Las tumbas de cámara megalíticas de España y Portugal se fechan entre 3800 y 2000 a.C. Las tumbas británicas e irlandesas datan de poco después del 3800 al 2000 a.C.; y las tumbas escandinavas, de antes del 3000 al 1800 aproximadamente. En la secuencia irlandesa conviene tener una fecha precisa para Newgrange, señalando que fue erigido alrededor de 2500 a.C. En la secuencia británica Atkinson ha consignado los detalles cronológicos de Stonehenge: la primera fase fue del 2800 al 2200 a.C.; la segunda fase, incluida la llegada de las piedras azules de Gales, se extendió del 2100 al 2000 a.C.; la tercera fase, que comprendió tres subfases, del 2000 al 1100 a.C., fecha en que comenzó la fase final.

Las fechas que he dado aquí, para Stonehenge y para los monumentos megalíticos en general, son lo que se llama fechas calibradas del carbono 14. Es decir, que han sido acomodadas a las correcciones fundadas en el estudio de los anillos del pino de piña erizada, conforme a las curvas de variación establecidas por Hans E. Suess, de la Universidad de California en San Diego, R. Malcolm Clark, de la Universidad de Monash, en Australia, y por otros investigadores.

¿Cómo se puede explicar el origen de estos monumentos megalíticos ahora que están exactamente fechados? Sólo hay una explicación fidedigna para los menhires de la Bretaña francesa y de otras partes del noroeste de Europa. Son una invención local; quizá representen el traslado a la piedra de una costumbre anterior de erigir postes de madera como cenotafios o lindes territoriales o incluso postes totémicos.

Ahora hay tan sólo una explicación para los templos megalíticos de Malta. Solía sostenerse que se derivaban, por difusión, de la Creta minoica o de Micenas, o de más lejos incluso. La cronología del carbono 14 demuestra actualmente que son más antiguos que cualquiera de los minoicos, micénicos, egipcios o sumerios. Los templos malteses fueron una invención indígena. Es posible que sean versiones sobre el suelo de templos-tumbas rupestres subterráneos. De todos modos, no parecen tener antecedentes en ninguna parte, y tampoco hay edificios que se puedan derivar confiadamente de los templos malteses en parte alguna fuera de Malta y de Gozo.

Llegamos ahora a los orígenes de las cámaras megalíticas y de los anillos de piedra. No han de contemplarse como fenómenos aislados. Debemos tratar más bien de explicarlos en el contexto de las sociedades neolíticas que los crearon. La idea de Elliot Smith de que los megalitos europeos procedían de las mastabas, o tumbas de piedra, de Egipto, carece de fundamento real y no tiene visos de probabilidad. Ahora sabemos que los megalitos de Europa son más antiguos que las mastabas y las pirámides de Egipto.

La nueva conclusión

Una hipótesis distinta, expuesta por V. Gordon Childe, entre otros, y sostenida tenazmente por la mayoría de los arqueólogos durante el segundo cuarto de este siglo, suponía que los megalitos de Europa habían sido construidos por gentes originarias del Mediterráneo oriental, particularmente de Creta y de las islas del Egeo. Pero aun en los días anteriores al carbono 14 se iba abriendo camino la idea de que derivar, por ejemplo, las grandes tumbas de cámara como Newgrange y Maes Howe de las tumbas abovedadas subterráneas de Micenas era imposible cronológicamente. Como segunda línea de defensa fue adoptada la hipótesis de remontarse en el tiempo a las tumbas abovedadas de Messará, en Creta, y a las tumbas rupestres de las Cícladas. Ahora las precisiones del carbono 14 y de la termoluminiscencia indican que las cámaras megalíticas europeas son anteriores a cualesquiera otras tumbas colectivas del Mediterráneo oriental.



EL MAYOR MENHIR de Europa occidental es este ejemplar roto de Locmariaquer, Bretaña. Se ven aquí tres de sus cuatro fragmentos; la longitud total de los cuatro es de 22 metros. No se sabe si el menhir llegó a estar erguido; los catálogos más antiguos lo muestran en el mismo estado en que se encuentra.

Nos vemos, pues, obligados a concluir que las tumbas de cámara se originaron independientemente en siete zonas de Europa por lo menos: sur de España, Portugal, Bretaña, norte de Francia, norte de Alemania, Escandinavia, sur de Inglaterra y Escocia. ¿Parece esta afirmación una victoria de los partidarios de la "evolución independiente", dentro del campo de la teoría arqueológica, sobre la escuela "difusionista" que durante tanto tiempo se mantuvo dueña del campo?

No es así de sencillo. Las sociedades neolíticas que comenzaron a levantar megalitos en determinadas áreas de Europa alrededor del 4000 a.C. fueron las que estaban construyendo ya edificios comparables no líticos o no megalíticos, es decir, casas y tumbas. En el momento actual somos muchos los que pensamos que la explicación del nacimiento de las tumbas de cámara megalíticas en muchas y distintas partes de Europa debe tener en cuenta tres fases sucesivas. La primera fase postula una tradición primitiva neolítica europea de construcción de casas de madera y piedra. La segunda fase conlleva la transformación de estos edificios domésticos en tumbas, construidas aún de madera, terrones o piedra no megalítica. La tercera fase supone la transformación de estas extensas tradiciones neolíticas en arquitectura megalítica en puntos de Europa distintos, aunque interrelacionados sin duda alguna.

El origen de los anillos de piedra puede explicarse en gran parte del mismo modo, si bien aquí el fenómeno no afecta a siete zonas separadas e interconectadas. Los anillos sólo se encuentran en Gran Bretaña y en Irlanda. Del mismo modo que la tumba de cámara traduce a un precursor más sencillo, los anillos de piedra no se pueden disociar de los monumentos en forma de anillos de madera a los que solemos referirnos como empalizadas (*woodhenges*).

Así veo yo el origen de los anillos de piedra. Primero hubo claros circulares en los bosques que cubrían la Europa neolítica en los milenios quinto y cuarto a.C. Podemos suponer que en estos claros se celebrasen reuniones sagradas y profanas. Seguidamente, debido a la acción de los animales domesticados por el hombre y a la del hombre mismo, los bosques desaparecieron, momento en que se formaron claros artificiales poniendo una empalizada de postes de madera en forma de anillo para reuniones del mismo tipo. La tercera fase fue la traducción de los anillos de madera a los anillos de piedra. Por último, como *tour de force*



MENHIR CRISTIANIZADO en la aldea de Rudston, Yorkshire. Con sus tres metros de altura es el más alto de Inglaterra. Este megalito no fue derribado. A su alrededor se fue instalando el camposanto.

de los que se podrían llamar arquitectos de catedrales, Stonehenge fue construido a mediados del tercer milenio a.C., y floreció como templo y al mismo tiempo como lugar de reunión y como estadio durante más de 1000 años. Esto nos hace retroceder a mi anterior pregunta: ¿Se adapta Stonehenge a la descripción de los anillos de piedra megalíticos, en general, en tanto que lugares de reunión sagrados y profanos? La respuesta parece que ha de ser decididamente afirmativa.

Megalitos y religión

Tenemos la sensación de que debió haber existido una fe impulsora que movió a la gente neolítica a trabajar con todas sus fuerzas para construir los megalitos de Europa. Las tumbas, los templos, los cenotafios, lugares de reunión, cualquier cosa que los monumentos fuesen, parecen manifestaciones de una profunda fe religiosa que comportaba una creencia en la vida de ultratumba. ¿Estaba la religión conectada con el ciclo anual del sol, como dan a entender la orientación de Stonehenge y de los otros anillos de piedra? ¿Estaba relacionada con una diosa madre como parecen apuntar las figuras de los templos de Malta y las de la cerámica y las placas de pizarra de las tumbas de España y de Portugal? Tal vez no lo sepamos nunca. Sin fuentes escritas es prácticamente imposible reconstruir los caracteres religiosos y sociales de las comunidades prehistóricas. No sabemos por qué los cazadores del Paleolíti-

co Superior crearon las pinturas, grabados y esculturas del sur de Francia y del norte de España, y debemos admitir que tampoco conocemos el ethos social y religioso de los constructores de megalitos.

Entre los miles de tumbas megalíticas de Europa hay unas cuantas decoradas con motivos grabados en bajorrelieve o incisos en la superficie de la piedra. Algunos de Alemania central parecen representaciones de tapices, quizá de las colgaduras que pueden haberse alineado en las paredes de las casas. Otros, particularmente en la cuenca de París, son claras representaciones de un personaje femenino visto de frente; los ojos, la nariz y los senos de la divinidad están realzados. La mayoría de los motivos megalíticos de las tumbas de Bretaña e Irlanda son, sin embargo, temas geométricos: espirales, zigzags, rombos, círculos concéntricos. Quizá los más hermosos se encuentren en Gavrinis, en Bretaña, y en Newgrange. Algunos, como las espirales de la gran piedra a la entrada de Newgrange, parecen puramente decorativos. Otros se muestran ante nuestros ojos como una maraña desconcertante de lo que pueden ser signos y símbolos. Hay que tener presente que algunas de estas decoraciones geométricas están enterradas en los túmulos que rodean a las tumbas de cámara y nunca se destinaron a ser vistas; quedaron ocultas deliberadamente desde el momento de la construcción ¿Por qué? ¿Son mensajes o son símbolos sagrados? Cualquiera que sea su significado, está transcrito en una notación que nunca llegaremos a descifrar.

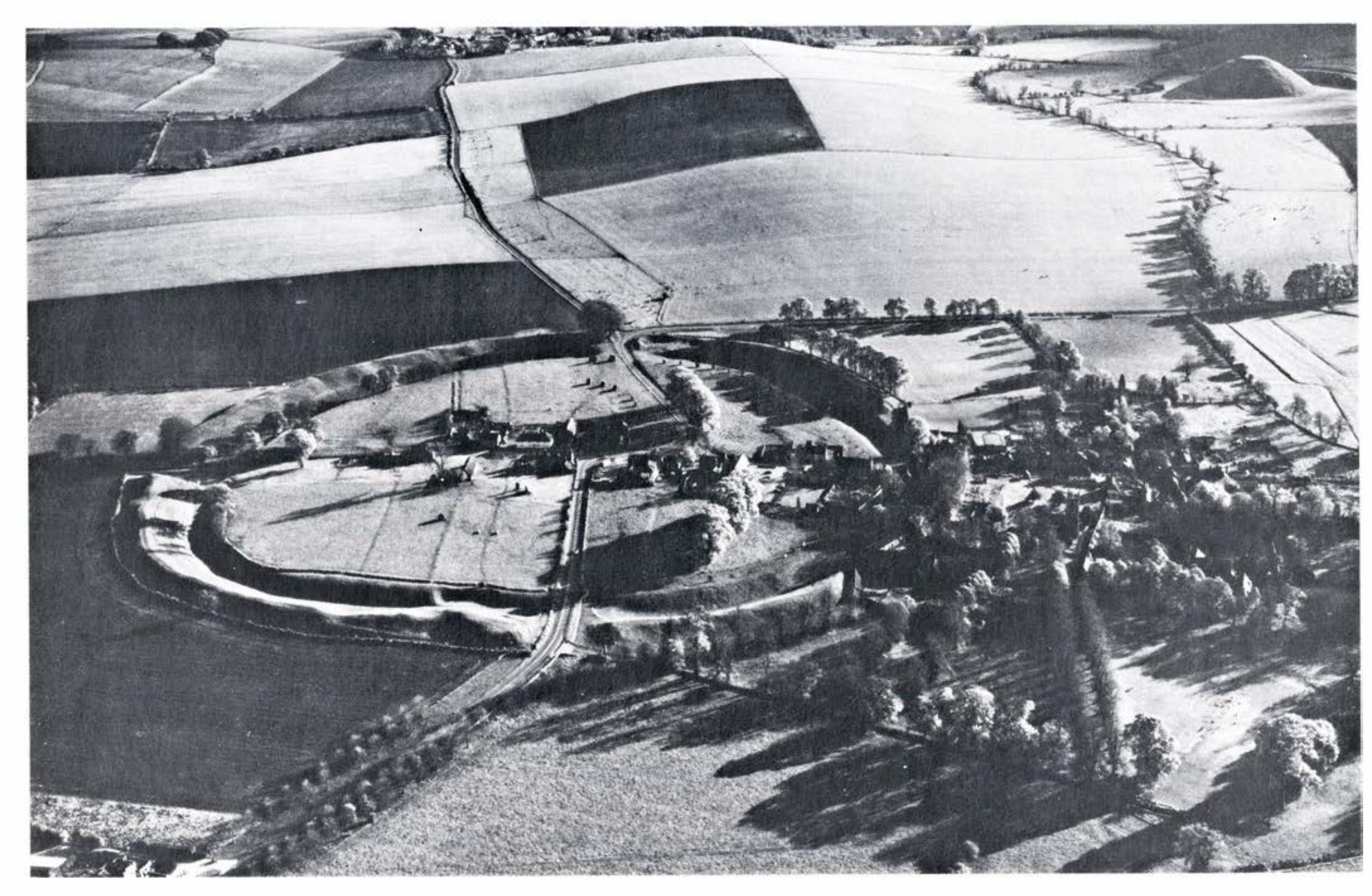
En relación con esto, se descubrió hace sólo quince años que la representación de un puñal había sido grabada en uno de los pies derechos de Stonehenge, como lo habían sido las figuras de varias hachas del estilo de la Primera Edad de Bronce. Algunos imaginaron que el puñal era micénico. El arte de este tipo es totalmente distinto del de las tumbas de cámara, y como es natural, estas representaciones pueden haber sido hechas en un momento tardío de Stonehenge. El decir que son "graffiti" de la Edad de Bronce no significa, sin embargo, que Stonehenge haya sido un gigantesco retrete de aquella época.

La costumbre de construir grandes monumentos de piedra toca a su fin en Europa hacia el año 1000 a.C., pero no hay que suponer que el pueblo en general olvidase el carácter y el significado de los edificios. No es en modo alguno imposible que la gente del primer milenio a.C. siguiese congregándose en los anillos de piedra adorándolos, y es más que posible que el sacerdocio druídico de los celtas prerromanos de la Galia y de Britannia los usase como templos. No hay, empero, prueba alguna arqueológica de ello.

En un primer momento el cristianismo se opuso terminantemente a que la gente diese culto a las piedras; pero poco a poco se abrió paso una nueva tolerancia, lo bastante generosa para cristianizar a algunos menhires. Más aún, en España y en Bretaña algunos monumentos megalíticos han sido incorporados a iglesias cristianas modernas y con culto. Yo interpreto esto como un signo de que la fe de los constructores sobrevivió de algún modo o manera hasta por lo menos la Edad Media de Europa occidental.

La pericia de los constructores

Antes era costumbre preguntar: ¿Quiénes son los constructores megalíticos de la Europa antigua y dónde

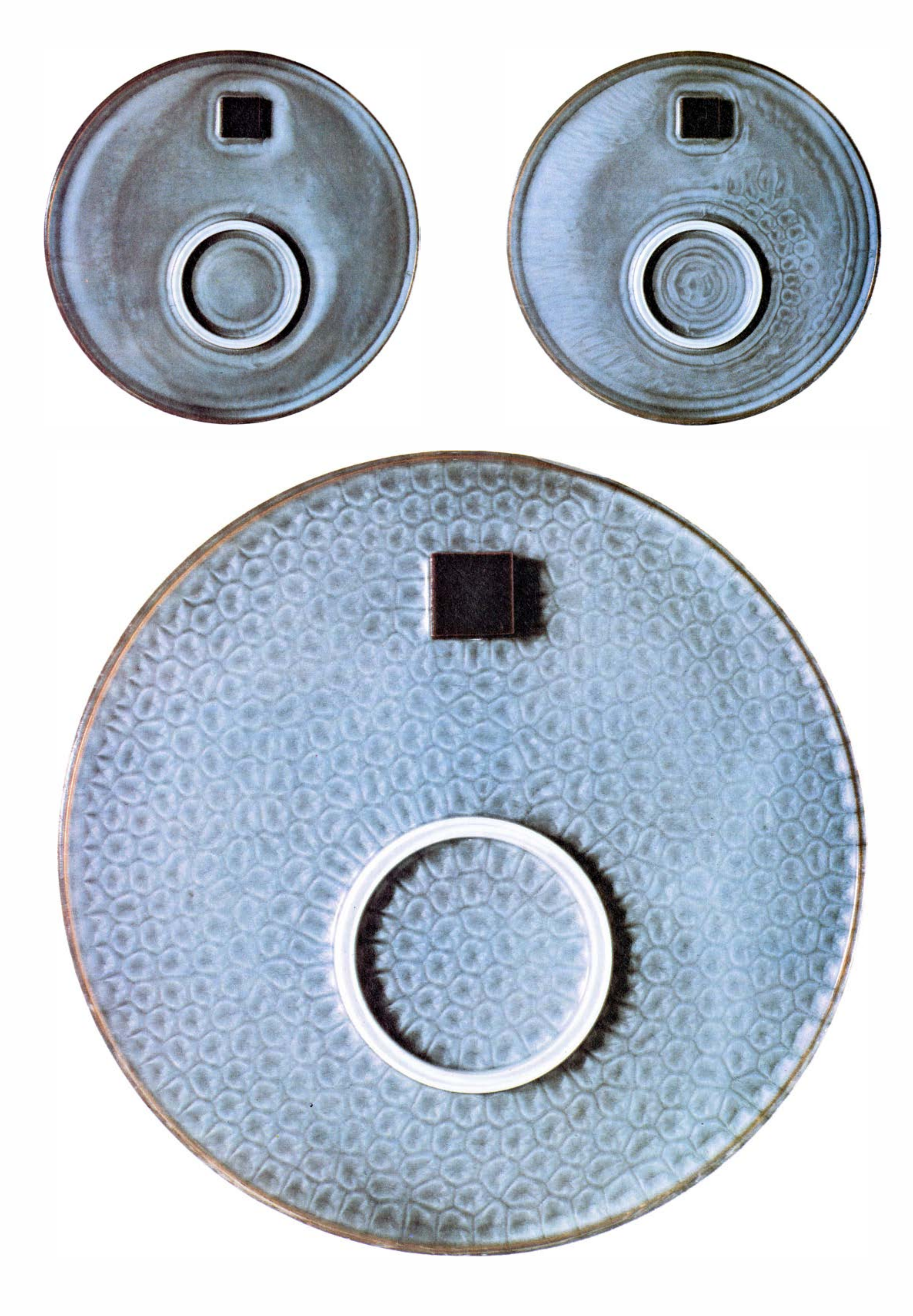


RECINTO DE AVEBURY, señalado por vez primera como círculo megalítico por John Aubrey en el siglo XVII, del que apuntó "sobrepasa tanto en tamaño al renombrado Stonehenge, como una catedral lo hace con respecto a una iglesia parroquial". Pocos de los 190 y pico menhires (o piedras erguidas) que componen el círculo y sus avenidas son visibles en esta fotografía aérea, pero el terraplén y el foso, formados por la extracción de 200.000 toneladas de caliza, delimitan una superficie de más de 14 hectáreas. La loma cónica visible a la derecha, llamada Silbury Hill, es también obra del hombre.

están sus casas y poblados? Ahora pensamos que esas preguntas están formuladas al revés. Fueron los aldeanos neolíticos de Europa los que en el cuarto y tercer milenio a.C. construyeron sus tumbas y templos de duradera piedra en ciertas zonas. Ultimamente se ha hablado y se ha escrito mucho acerca de una revolución en nuestra imagen del pasado prehistórico. Se imputa a los arqueólogos haber concebido a los pueblos neolíticos como salvajes, mientras que la nueva visión del pasado revela que poseían una gran pericia matemática, geométrica y astronómica, lo que hace cambiar de forma espectacular nuestra imagen del mundo prehistórico.

Esta tesis sólo tiene sentido para quienes nunca han entendido los registros arqueológicos y quieren hacer de la prehistoria algo sensacional. Yo nunca he dudado de que los pueblos neolíticos de Europa eran buenos técnicos y expertos ingenieros. Extraían grandes piedras, las transportaban a grandes distancias y las erigían con extremada pericia y sentido artístico. Podemos aprender algo acerca de sus probables técnicas mediante el estudio de los métodos que hoy se emplean en zonas como Assam, donde todavía se construyen monumentos megalíticos. Podemos calibrar sus realizaciones observando las dificultades que los granjeros y constructores actuales encuentran para demoler los edificios prehistóricos. Por ejemplo, a fines del siglo pasado la clave de una cámara megalítica cercana a Saumur fue trasladada para aprovecharla como puente sobre un arroyo de la localidad. Los transportistas construyeron una serie de enormes rodillos de más de un metro de perímetro; cada rodillo estaba hecho amarrando los troncos de cuatro robles. Después de emplazados los rodillos, hicieron falta dieciocho parejas de bueyes para transportar la carga.

No debemos negar la grandeza de las realizaciones de quienes erigieron los megalitos. Ni deberíamos negar que de vez en cuando, haciendo una pausa en sus faenas de la cosecha o de la construcción de un monumento, dirigiesen la vista, como todos hacemos, hacia el sol, la luna y las estrellas. Hace 6000 años que se construyeron los primeros megalitos en Malta y en Bretaña, 4000 años antes de la Era Cristiana y 1000 antes de las civilizaciones literarias de Egipto y Mesopotamia. Es este un pensamiento reconfortante cuando contemplamos con orgullo y con placer nuestro patrimonio megalítico.



Convección

Sólo será posible entender las corrientes ascendentes y descendentes en un fluido calentado cuando se describan las complejas relaciones que existen entre temperatura, viscosidad, tensión superficial y demás propiedades del fluido

Manuel García Velarde y Christiane Normand

La convección, movimiento de un fluido que sirve para el transporte de masa o calor, debe ser un fenómeno corriente para todo el que haya observado la turbidez de un caldo caliente (provocada por la agitación de las partículas suspendidas en él), se haya fijado en cómo asciende el humo que sale por una chimenea o haya notado el reflejo de la luz por las corrientes de aire que se forman sobre una carretera asfaltada en un día caluroso. Idéntico mecanismo de corriente convectiva provoca las grandes corrientes oceánicas, la circulación global de la atmósfera y, a una escala aún mayor, origina los movimientos sobre la fotosfera solar. Asimismo, algunas nubes se forman cuando aire húmedo caliente asciende por convección; y por otro lado, los altos niveles de contaminación que se alcanzan periódicamente sobre las grandes ciudades, como Madrid o Los Angeles, por una inversión de temperaturas, se deben bien a la ausencia de transporte convectivo, o bien a la existencia de un movimiento convectivo térmico en trayectorias cerradas que impide la difusión de las partículas suspendidas hacia capas más altas de la atmósfera. Hay muchos otros ejemplos menos corrientes o más difícilmente observables. Así, la convección tiene una influencia importante en el secado de películas de pintura y en la difusión de gases y partículas en los pulmones. Igualmente, la lenta deriva de los continentes parece ser debida a la convección del manto terrestre.

El caso más sencillo de convección aparece cuando un fluido (un gas o un líquido) se calienta por abajo. Debido a este calentamiento, la capa inferior del fluido se expande y se hace menos densa que las capas superiores. En esta situación, la capa inferior más caliente y ligera tenderá a elevarse y la capa superior más fría y densa a caer. A primera vista, parece que esta sencilla explicación bastaría para los tipos más elementales de convección. Sin embargo, aún hoy día, no disponemos de una rigurosa y exacta descripción matemática para el sistema más simple que experimente un intenso movimiento convectivo y la formulación de la teoría detallada y cuantitativa de la convección sigue siendo un reto científico.

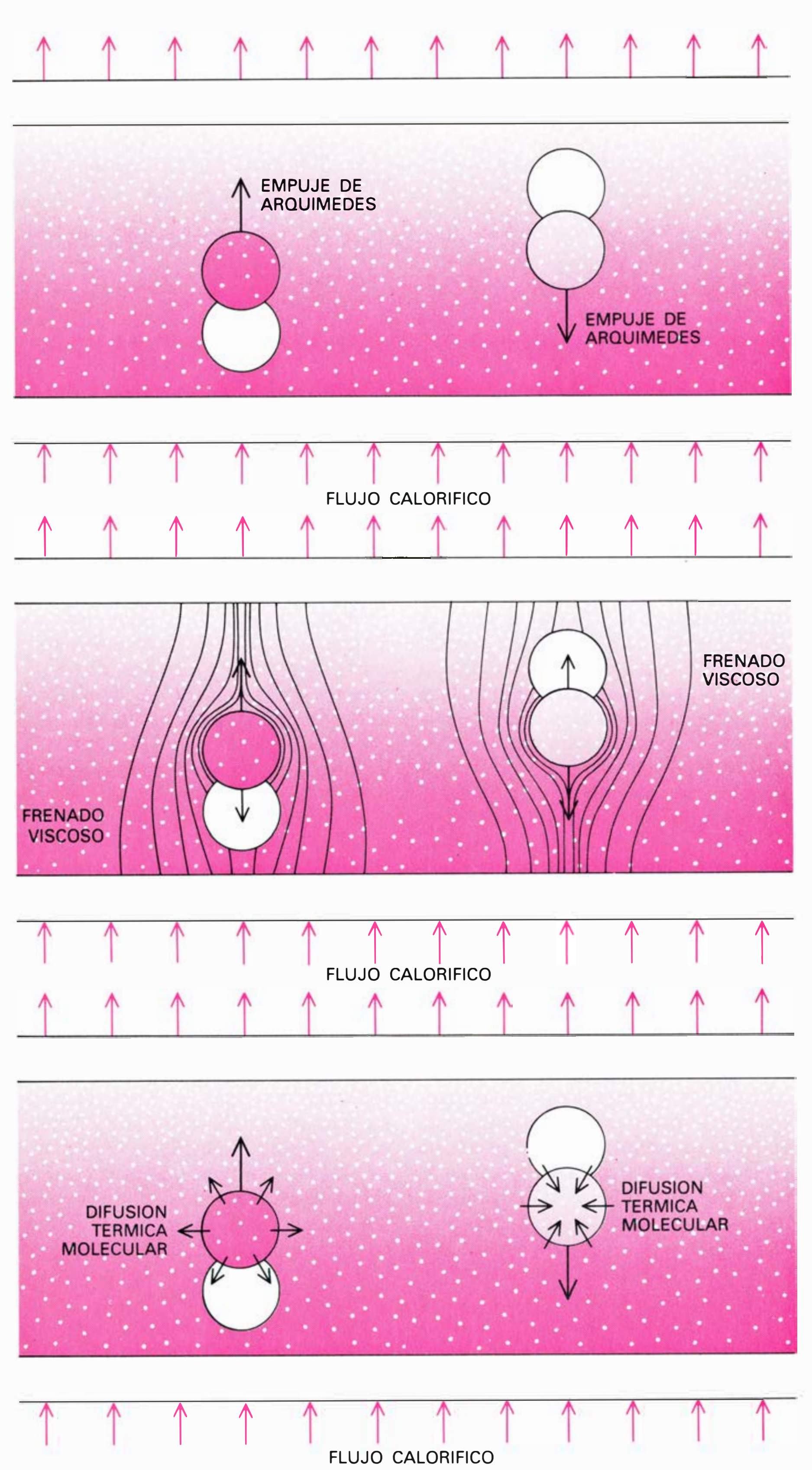
Las dificultades teóricas inherentes al problema pueden comprenderse, considerando, de nuevo, el caso de una capa de fluido calentada por abajo. En dicho fluido, el empuje de Arquímedes es la fuerza que origina la convección y la magnitud de esta fuerza viene determinada por la diferencia de temperatura entre la parte superior e inferior de la capa. La complejidad del problema se hace evidente cuando nos damos cuenta de que la distribución de temperaturas se ve alterada por la corriente convectiva que lleva calor desde la parte inferior a la superior de la capa; de modo que la fuerza que origina la corriente se ve modificada por dicha corriente.

Pese a que no hay solución rigurosa a este problema, en las últimas dos décadas ha habido indudables progresos hacia una teoría general de la convección. Los avances se han debido, principalmente, al haberse adoptado ideas y técnicas matemáticas desarrolladas en otros campos de la física; tales como el estudio de las transiciones de fase de materiales ferromagnéticos y de la superconductividad. Con tales métodos matemáticos se hace un análisis de estabilidad de los varios posibles movimientos en el fluido y, de ahí, se obtienen predicciones de los que cabe observar. Pese a ser sólo aproximados, los resultados a veces son muy aceptables y útiles. Tenemos, pues, cierta esperanza de que en el futuro se pueda explicar rigurosamente lo que ocurre en un sopicaldo.

El transporte convectivo que discutiremos aquí se denomina convección libre o natural, indicando que la corriente se debe a fuerzas que actúan en el interior del fluido. En la mayoría de los casos dicha fuerza es la gravedad, aunque bajo determinadas circunstancias algún otro agente, como la tensión superficial o un campo electromagnético, juegue un papel relevante e incluso predominante. Se habla de convección natural para distinguirla de la convección forzada, en la cual el movimiento del fluido se debe a una fuerza impuesta desde el exterior, por ejemplo, el gradiente de presión creado por un compresor o un ventilador.

Benjamin Thompson, Conde de Rumford, en la última década del siglo XVIII introdujo la idea de la convección natural para describir el transporte calorífico observado en una tarta de manzana. Posteriormente hubo varios intentos de relacionar los mecanismos convectivos con la circulación atmosférica y a lo largo de todo el siglo XIX se fueron añadiendo a la literatura un gran número de informes. Sin embargo, hasta los albores del año 1900 no se llevaron a cabo investigaciones sistemáticas. Fue crucial en esa época el trabajo experimental del científico francés Henri Bénard. Bénard estudió

CELDAS CONVECTIVAS con una característica geométrica poligonal; éstas aparecen espontáneamente cuando una fina capa de fluido, con una superficie libre, se calienta por abajo. En las fotografías de la página anterior se sigue la evolución temporal de la convección; la fotografía inferior muestra la estructura de la corriente convectiva totalmente desarrollada. Inicialmente, las celdas son "rollos" alargados cuya forma empieza siendo impuesta por los contornos laterales de la capa; posteriormente, los rollos se transforman en polígonos, en principio irregulares, que tienden a una forma de hexágono regular independientemente de la geometría de la capa (sus contornos). En cada celda, el líquido asciende por el centro y cae por sus laterales (al contrario, en el caso de gases). Este tipo de circulación se obtiene cuando la tensión superficial juega un papel predominante en el desarrollo de la corriente y, generalmente, sólo se establece esta estructura poligonal cuando la superficie superior es libre. Las fotografías se tomaron en el laboratorio de Velarde, en la Universidad Autónoma de Madrid. El fluido era un aceite de silicona con limaduras de aluminio suspendidas en él para visualizar la corriente.



CUANDO LAS FUERZAS que actúan sobre una región de fluido no están en equilibrio se inicia la convección. Dichas fuerzas pueden analizarse en un experimento ideal en que una fina capa de fluido entre dos placas rígidas horizontales se calienta por abajo; debido a este calentamiento, se generan en el fluido gradientes verticales de temperatura y densidad. Si una región de fluido caliente, en las proximidades de la placa inferior, se desplaza levemente hacia arriba, se introduce en una región de mayor densidad media, estando, por tanto, la región desplazada sometida a un empuje ascendente (principio de Arquímedes). Análogamente, si una región de fluido frío, en las proximidades de la placa superior, se desplaza hacia abajo, tenderá a seguir cayendo, puesto que será más pesada que sus alrededores. El frenado viscoso y la difusión calorífica molecular se oponen al empuje de flotación. La importancia relativa de estas fuerzas se mide por una relación adimensional denominada número de Rayleigh. La convección se inicia cuando el empuje supera los efectos disipativos del frenado viscoso y la difusión calorífica; es decir, cuando el número de Rayleigh sobrepasa un cierto valor crítico.

un sistema convectivo, más complicado de lo que pensaba, cuyo intríngulis no ha sido reconocido hasta hace bien pocos años. Más adelante se describen las observaciones de Bénard y su actual interpretación.

El gran científico de la convección natural, al comienzo del siglo XX, fue John William Strutt, Lord Rayleigh, quien intentó obtener teóricamente los resultados experimentales de Bénard en uno de sus últimos trabajos, publicado en 1916. Hoy se sabe que la teoría de Lord Rayleigh no es aplicable al sistema estudiado por Bénard; a pesar de lo cual, el trabajo de Lord Rayleigh es el punto de partida de casi todas las teorías actuales sobre convección natural.

La teoría de Lord Rayleigh puede explicarse en el contexto de un experimento ideal con un fluido cuyas propiedades sean mucho más simples que las de cualquier líquido o gas real. Una fina capa de este fluido ocupa totalmente el espacio entre dos placas planas, rígidas y horizontales, de modo que no exista ninguna superficie libre. Por capa fina se entiende una cuyas dimensiones horizontales sean mucho mayores que su espesor (siendo éste, de hecho, igual a la separación entre las placas). Considerar la capa como fina se hace para minimizar la influencia de las paredes laterales del recipiente, y así no tenerlas en cuenta, explícitamente, en la descripción teórica. Teóricamente, la capa debería ser de extensión horizontal infinita; en la práctica, una capa de varios centímetros de longitud y pocos milímetros de espesor verifica estas condiciones.

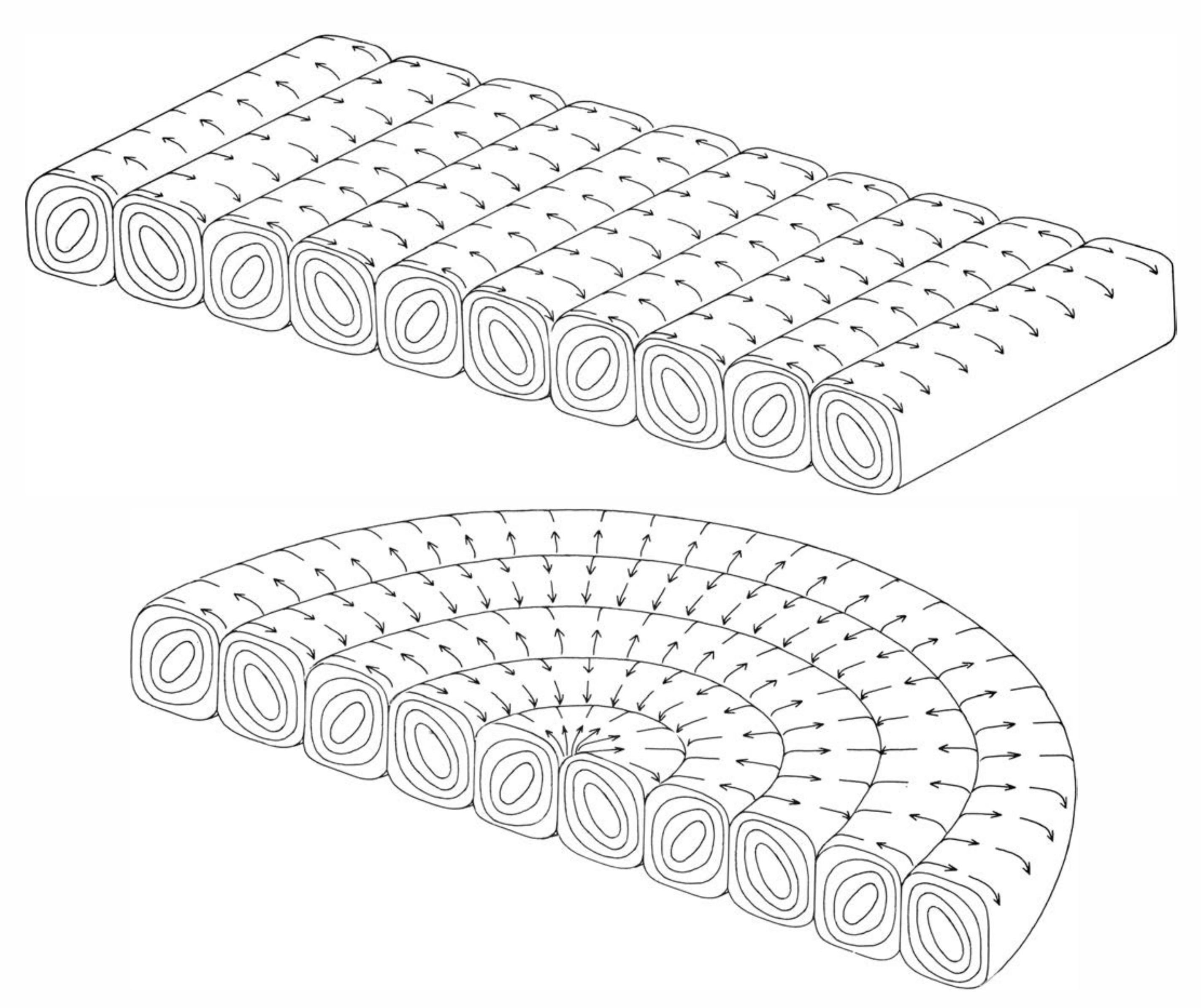
El sistema se calienta por abajo, de forma que la placa inferior del recipiente se mantiene a una temperatura constante y uniforme. Análogamente, se extrae calor por la placa superior, de forma que la temperatura, también aquí, se mantiene constante y uniforme. Por tanto, la diferencia de temperatura entre las dos placas es constante y la temperatura será la misma para todos los puntos del fluido que se encuentren en un mismo plano horizontal. En consecuencia, el gradiente de temperatura (la variación de la temperatura con la posición vertical) será lineal; es decir, si se representa la temperatura en función de la altura se obtiene una línea recta.

Aparte de éstas (el gradiente de temperatura es lineal y las paredes laterales no influyen sobre los movimientos del fluido) cabe hacer otras hipótesis que simplifiquen aún más el tratamiento del problema. Así, se supone que la única fuerza que actúa en

el fluido es la gravedad. Como los experimentos de laboratorio son de pequeña escala, la aceleración de la gravedad se considera constante, ya que el campo gravitatorio será prácticamente uniforme en todo el volumen. Además, se supone que el fluido es incompresible (las diferencias de presión en el interior del fluido no afectan a la densidad); en el caso de líquidos, si el espesor no es muy grande, esta hipótesis resulta muy aceptable. Por último, se supone que únicamente una propiedad del fluido depende de la temperatura: la densidad, que disminuye cuando la temperatura aumenta. En otras palabras, el fluido se expande cuando se calienta; tal es lo que suele ocurrir en gases y líquidos reales. Esta última hipótesis es la más restrictiva de todas, ya que en fluidos reales hay más propiedades que dependen de la temperatura, como, por ejemplo, la viscosidad y el coeficiente de difusión térmica.

Con este experimento ideal, hay un camino teórico sencillo para investigar la influencia del gradiente sobre el estado inicial de reposo del fluido. Consiste en suponer que una región del fluido se ha desplazado verticalmente (hacia arriba o hacia abajo) de su posición de equilibrio y estudiar a continuación las fuerzas que actúan sobre el fluido. Para una primera aproximación, el desplazamiento tiene que ser pequeño (formalmente, la teoría de Rayleigh es únicamente válida para desplazamientos infinitesimales), aunque la región desplazada puede tener cualquier forma o tamaño. Cabe imaginar que dicho desplazamiento inicial no esté causado por una fuerza exterior; ya que las moléculas del fluido están en agitación constante, sus posiciones varían aleatoriamente y es de esperar que tal desplazamiento infinitesimal ocurra espontáneamente en el transcurso del tiempo.

Consideremos, pues, una región de fluido en las proximidades de la placa inferior. Debido a la mayor temperatura en esa zona, dicha región tiene una densidad menor que la densidad media de la capa. Sin embargo, siempre que permanezca en el lugar que ocupa, estará rodeada de fluido de la misma densidad y, por tanto, no estará sometida a ningún empuje. Todas las fuerzas que actúan sobre dicha región están en equilibrio, de modo que ésta no realizará ningún movimiento.



CELDAS EN FORMA DE ROLLOS son la configuración estable en la convección debida al empuje de flotación. Dos rollos adyacentes que rotan en sentidos opuestos forman la unidad básica de la configuración; la anchura de esta unidad es dos veces el espesor de la capa de fluido. La geometría y orientación de estos rollos depende de la forma de las paredes laterales. Así, en un recipiente de base rectangular, los rollos tienden a colocarse paralelos al lado menor, mientras que, en un recipiente circular, los rollos forman roscos concéntricos. En la convección dirigida por tensión superficial, esta estructura de rollos no es estable; de modo que una configuración estable de rollos se observa sólo cuando el fluido no tiene ninguna superficie libre.

Supongamos que, por cualquier perturbación interna, la región de fluido se desplaza levemente hacia arriba. ¿Seguirán estando en equilibrio las fuerzas que actúan sobre ella? Ahora, la región de fluido está rodeada por fluido más frío y, por tanto, más denso. Por lo cual, actúa sobre ella un empuje de Arquímedes que favorece el movimiento ascendente. La fuerza neta hacia arriba es esencialmente proporcional a la diferencia de densidades y al volumen de la región desplazada. Así, el gradiente de densidades amplifica el desplazamiento inicial, ya que la fuerza resultante favorece la continuación del movimiento. Puede realizarse un análisis similar para una región de un fluido frío y denso en las proximidades de la plaza superior que sufre un leve desplazamiento vertical hacia abajo. En este caso, se introduce en una zona de menor densidad media, siendo la región desplazada más pesada que sus alrededores de forma que tenderá a continuar su movimiento descendente, amplificándose también la perturbación inicial. La convección natural es el resultado de la combinación de estas corrientes ascendentes y descendentes, que tienden a cubrir toda la capa de fluido, aunque en cualquier experimento suele iniciarse junto a las paredes laterales.

El análisis, desarrollado hasta aquí, implica que la convección debe observarse en el fluido para cualquier gradiente de temperatura, independientemente de su valor numérico. Incluso para un gradiente infinitesimal, cualquier movimiento ascendente de fluido caliente o descendente de fluido frío debería bastar para el establecimiento de la corriente. Sin embargo, tal caso no es lo que se observa, sino que se precisa que el gradiente de temperatura supere un cierto valor crítico para que se desarrolle el movimiento convectivo. Una contribución muy importante de Lord Rayleigh fue mostrar la existencia de tal valor crítico.

Lord Rayleigh indicó que una teoría de la convección debe tener en cuenta, al menos, otros dos mecanismos que influyen sobre el movimiento del fluido. Uno de ellos, el frenado viscoso, equivalente en un fluido al rozamiento entre sólidos. Las fuerzas viscosas se oponen siempre al movimiento del fluido; estando, en parte, determinada su magnitud por una propiedad intrínseca del fluido, la viscosidad de cizalladura, que mide la resistencia al movimiento relativo entre dos regiones adyacentes de fluido. Si la velocidad no es demasiado grande, la magnitud de la fuerza viscosa es proporcional a la viscosidad multiplicada por el radio de la región de fluido y la velocidad. Obviamente, si el frenado viscoso es igual al empuje de Arquímedes, no puede haber movimiento.

Por otra parte, la convección no es el único medio de transporte calorífico en un fluido. También puede transportarse energía por radiación o por conducción (difusión molecular de calor). Para las bajas temperaturas en que se realizan la mayoría de los experimentos sobre convección, la radiación tiene una contribución poco significativa sobre el flujo calorífico, de modo que se pueden despreciar sus efectos en este problema. En cambio, la difusión molecular de calor nunca es despreciable y tiende a anular el gradiente de temperatura que originalmente potencia la corriente convectiva.

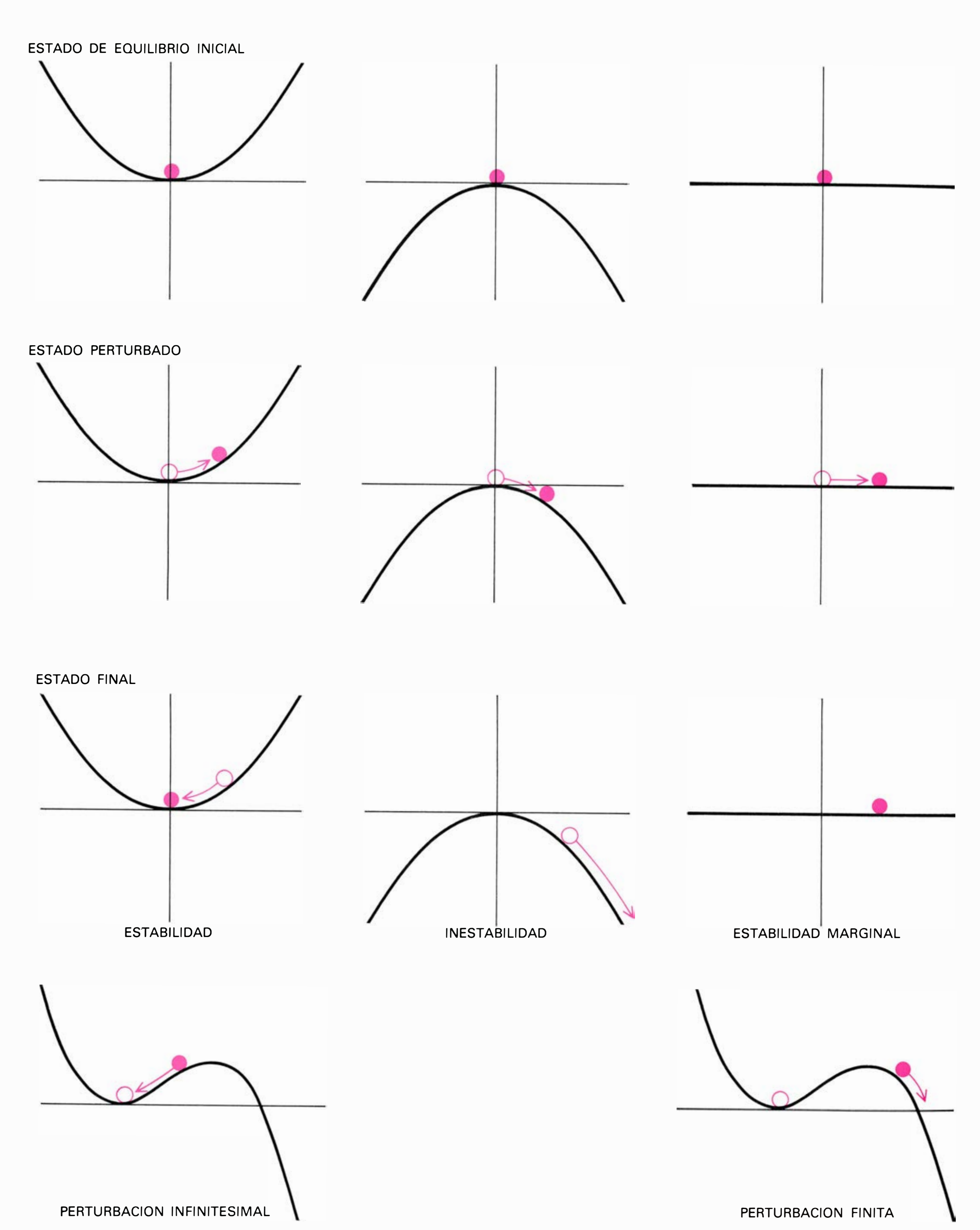
El efecto de la difusión de calor puede explicarse considerando, una vez más, una región de fluido caliente desplazada verticalmente hacia arriba, desde su posición de equilibrio hasta un entorno más frío. De acuerdo con la definición estadística de temperatura, las moléculas en la región más caliente tienen una velocidad media mayor que las que se encuentran en el fluido adyacente más frío. Las moléculas pueden cruzar libremente el contorno que delimita dicha región, y el efecto de estos intercambios de moléculas en ambos sentidos será igualar las velocidades medias (es decir, la temperatura) en las dos regiones. En otras palabras, el calor fluirá desde la región caliente hacia la región fría, de modo que la primera se enfriará y la segunda (los alrededores) se calentará hasta alcanzar el equilibrio térmico a una misma temperatura. Para una región de fluido desplazada verticalmente hacia abajo, el calor va en sentido opuesto: desde los alrededores más calientes hacia la región fría. En cualquier caso, la diferencia local de temperatura se verá reducida y, por ende, el empuje de Arquímedes inducido por aquélla.

El tiempo necesario para que una región de fluido alcance el equilibrio térmico con su entorno depende, principalmente, de una propiedad intrínseca del fluido: la difusividad térmica molecular. Su orden de magnitud en el tiempo es inversamente proporcional a la constante de difusividad y directamente proporcional al área superficial de la región considerada. Si este tiempo, característico del proceso de difusión térmica, es comparable al tiempo necesario para que la región de fluido se desplace una distancia dada, tal como su propio diámetro, desaparecerá dicho empuje o fuerza de flotación. Es decir, no podrá mantenerse una corriente convectiva si el fluido no se mueve más rápidamente que intercambia calor por difusión. El calor introducido en el sistema por la placa inferior será conducido a través de la capa de fluido por un mecanismo puramente conductivo o difusivo, sin que exista movimiento macroscópico en el fluido.

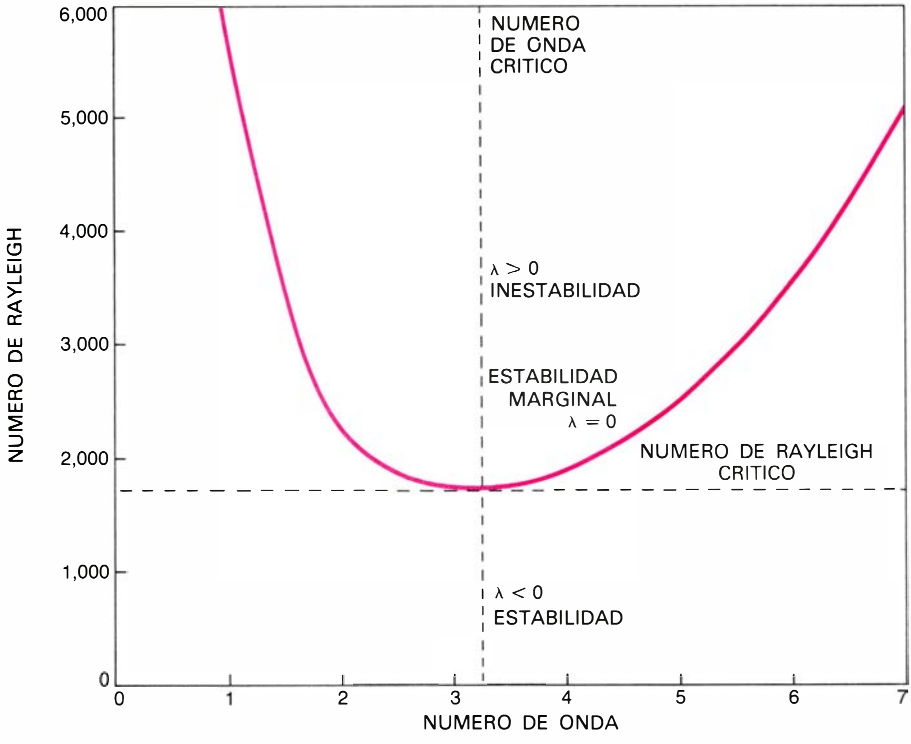
Con el análisis de Lord Rayleigh se demuestra que la mera existencia de un gradiente de temperatura no es suficiente para asegurar el inicio de la convección natural. Es necesario que el empuje, inducido por este gradiente de temperatura, supere los efectos disipativos de viscosidad y difusión de calor. La energía potencial gravitatoria liberada por la caída del fluido más denso y la elevación del fluido más ligero debe ser mayor que la energía disipada por viscosidad y por difusión calorífica. La relación entre estos efectos puede expresarse como un número sin dimensiones: el empuje de Arquímedes dividido por el producto del frenado viscoso y la difusión calorífica. Este número sin dimensiones, puesto que las unidades en que se miden numerador y denominador son las mismas, es una magnitud cuyo valor no depende del sistema de unidades elegido. Dicha magnitud se denomina número de Rayleigh. La convección se origina cuando el número de Rayleigh sobrepasa un cierto valor crítico.

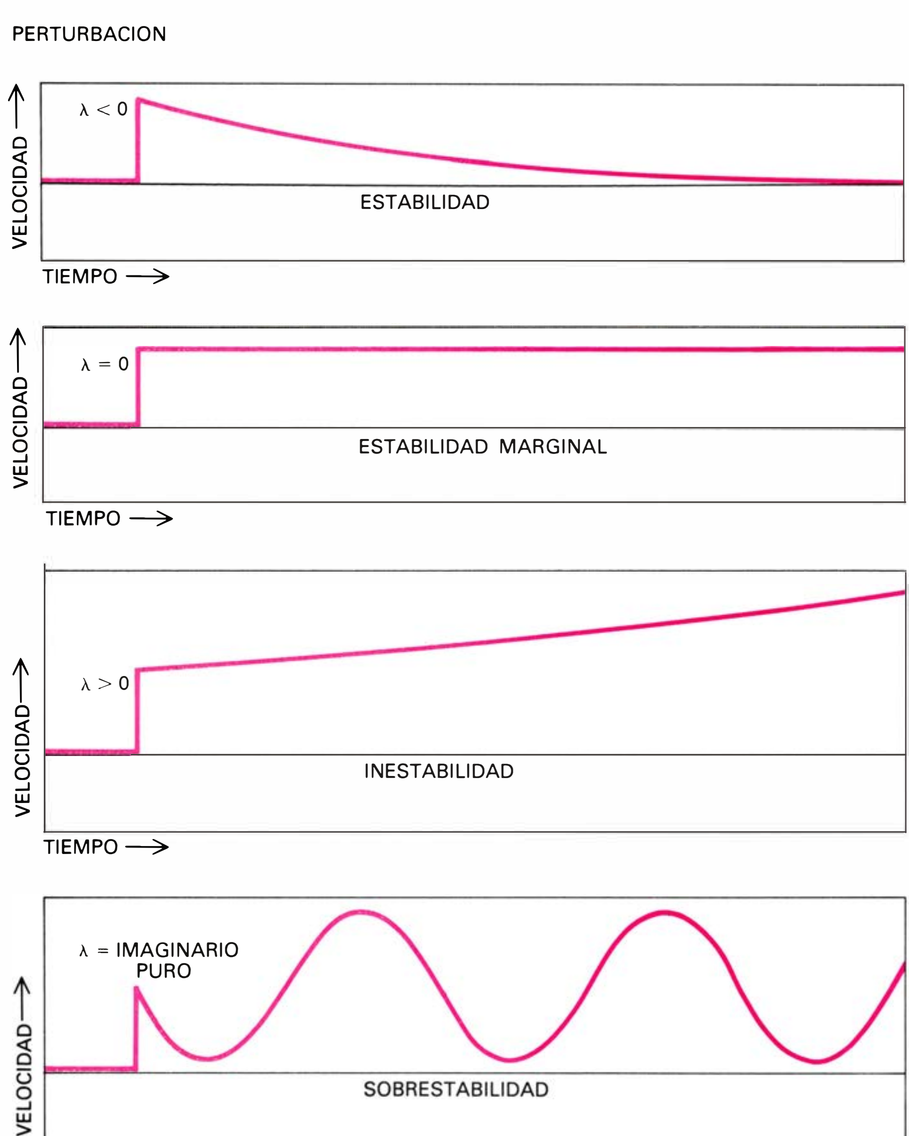
El significado del número de Rayleigh puede entenderse con mayor precisión a partir de un estudio de la estabilidad de los distintos movimientos que pueden desarrollarse en el fluido. Es conveniente definir la estabilidad en términos de una curva o superficie potencial que dé la energía de un estado del sistema en función de una cierta variable. El sistema se encuentra, normalmente, en el estado de energía mínima, correspondiente al punto más bajo sobre la superficie potencial.

Veamos una imagen ilustrativa de la superficie potencial. Supongamos un tazón semiesférico con una canica en su interior. En el equilibrio, la canica permanece en reposo en el fondo del tazón, donde su energía potencial gravitatoria es menor. Si, debido a cualquier perturbación, la canica se desplaza levemente de la posición de equilibrio, cae de nuevo hacia la posición inicial. Quizás en su caída sobrepase esta posición y comience a oscilar en torno a ella, pero los efectos disipativos



ESTABILIDAD de un sistema físico: se analiza estudiando su comportamiento frente a perturbaciones arbitrarias. Por ejemplo, en el caso de una canica en reposo sobre una superficie, dicha perturbación podría ser un pequeño desplazamiento de la canica sobre la superficie. Si la superficie es cóncava hacia arriba, la canica vuelve a su posición de equilibrio en el fondo; se dice entonces que la posición de la canica es una posición de equilibrio estable. También, la canica puede encontrarse en equilibrio en el vértice de una superficie convexa (cóncava hacia abajo), pero en este caso el equilibrio es inestable: cualquier perturbación por pequeña que sea se amplifica, ya que la canica disminuye su energía potencial rodando hacia abajo sobre la superficie. Sobre una superficie plana, la canica ni vuelve a su posición inicial ni se aleja de ella y se dice que el estado de la canica es de estabilidad neutra o marginal. En un fluido puro, una condición para el inicio de la convección es que exista una distribución inestable de alguna propiedad del fluido.





de fricción amortiguarán las oscilaciones de modo que, en un instante posterior, la canica se encontrará de nuevo en reposo en el punto de mínima energía. Puesto que la canica, en el fondo del tazón, vuelve a su posición inicial después de una perturbación, se dice que la canica se encuentra en un estado de equilibrio estable (matemáticamente hablando se dice que es asintóticamente estable).

Cabe colocar el tazón semiesférico boca abajo, poniendo cuidadosamente la canica en la cúspide. También éste es un estado de equilibrio, ya que todas las fuerzas que actúan sobre la canica están en equilibrio y, en ausencia de perturbaciones, la canica permanece indefinidamente en reposo. Sin embargo, cualquier influencia externa (una ráfaga de viento, un portazo, etcétera) destruirá este precario equilibrio. Tras la perturbación, por muy pequeña que sea, la canica no vuelve al punto de equilibrio, sino que se aleja del punto inicial. No importa cuán pequeña sea la perturbación inicial, la canica se encontrará, al cabo de cierto tiempo, a una gran distancia (grande comparada con la perturbación) de su posición inicial. Por tanto, la perturbación se amplifica y se dice que el estado de equilibrio es inestable.

Existe una tercera posibilidad: situar la canica sobre una superficie plana. En este caso, cuando la canica se desplaza, ni vuelve a su posición inicial ni continúa alejándose de ella, simplemente permanece en su nueva posición. Sobre una superficie potencial plana, cada punto representa un estado de estabilidad neutra o marginal. Con esta imagen gráfica cabe darse cuenta de que la estabilidad absoluta de un estado del sistema sólo puede demostrarse sometiéndole a todo tipo de posibles pertur-

EL VALOR del número de Rayleigh determina la estabilidad de un fluido en el caso de convección debida al empuje de Arquímedes. Si su valor es menor que un cierto valor crítico, cualquier perturbación en el seno del fluido se amortigua, mientras que si el número de Rayleigh sobrepasa el valor crítico, la perturbación se amplifica con el tiempo. Suele ocurrir que el sistema sea más sensible a perturbaciones con un número de onda característico correspondiente a una escala de longitud de dos veces el espesor de la capa de fluido. En la teoría de Rayleigh, la velocidad depende exponencialmente de un parámetro λ y del tiempo *t*, estando el valor de λ determinado por la diferencia entre el número de Rayleigh y el valor crítico de éste. Si λ es negativo, la velocidad tiende a decaer; si λ es igual a cero, la velocidad permanece constante y si λ es positivo la velocidad aumenta con el tiempo. En general, λ es un número complejo; cuando la parte real de λ es igual a cero, y su parte imaginaria diferente de cero, se da un fenómeno denominado sobrestabilidad, que puede conducir a la oscilación periódica de la velocidad (y demás propiedades del fluido) con el tiempo.

baciones y viendo su respuesta. Por ejemplo, una canica que esté en el interior del tazón volverá al centro, después de una perturbación infinitesimal e incluso después de una pequeña perturbación finita; pero no volverá al punto inicial si la perturbación es suficientemente grande como para hacer salir a la canica del tazón. Es difícil demostrar que un estado de equilibrio es estable; en cambio, la inestabilidad puede verse sin más que comprobar que existe una perturbación cualquiera, aún la más ligera, que se amplifica espontáneamente.

La aplicación de estas ideas al problema de la convección es bastante sencilla. Es fácil demostrar que una capa de fluido en reposo, calentada uniformemente por abajo, se encuentra en un estado de equilibrio, incluso aunque fuesen despreciables los efectos viscosos y la difusión calorífica. Aunque la zona inferior del fluido es más ligera que la superior, de forma que la energía potencial gravitatoria podría reducirse intercambiando estas posiciones, en ausencia de perturbaciones todas las fuerzas que actúan sobre una región dada del fluido están en equilibrio. Una teoría de la convección debe indicar si el equilibrio es estable, inestable o si el sistema se encuentra en un estado de estabilidad marginal. En otras palabras, la teoría debe definir la forma de la superficie potencial.

La curvatura de la superficie potencial viene determinada por el número de Rayleigh. Si el número de Rayleigh es cero (es decir, el gradiente de temperatura y, por tanto, el empuje de flotación son nulos), el estado de reposo será claramente estable y la superficie potencial será cóncava hacia arriba como el interior de un tazón. Cualquier movimiento del fluido correspondería a un estado con una energía mayor que la del estado de reposo. En cambio, si el número de Rayleigh es suficientemente grande, de modo que el empuje de flotación supera todos los efectos disipativos, el fluido podrá reducir su energía total desarrollando una corriente convectiva. Cualquier perturbación del estado de equilibrio tenderá a amplificarse. En este caso, la forma de la superficie potencial será cóncava hacia abajo, como un tazón invertido.

La continuidad del potencial exige que exista un valor intermedio del número de Rayleigh para el cual la fuerza de flotación y las fuerzas disipativas sean de la misma magnitud. Este valor es el número de Rayleigh crítico, que determina que la superficie potencial sea "plana" y tengamos estabilidad marginal. Según aumenta el número de Rayleigh desde cero (debido a un aumento del gradiente de temperatura), la superficie potencial reduce su concavidad y se aplana gradualmente; para el número de Rayleigh crítico, la superficie potencial es perfectamente plana; y, si se continúa aumentando el número de Rayleigh, la superficie se transforma en convexa. El estado de reposo no llega a ser un estado inestable hasta que no se sobrepasa el valor crítico. Para el experimento ideal referido anteriormente, los cálculos indican un valor para el número de Rayleigh crítico de 1708. En un experimento típico de laboratorio, siendo el fluido una capa de aceite de silicona de algunos milímetros de espesor, el número de Rayleigh crítico se alcanza cuando la diferencia de temperaturas entre placas es de pocos grados Celsius.

Aunque la teoría de Lord Rayleigh se base en un experimento hipotético que incluye muchas hipótesis simplificadoras, algunas de las cuales se sabe, de hecho, que no se verifican, es notablemente fructífera para predecir las condiciones suficientes para el inicio de la convección en fluidos reales. Por ejemplo, los experimentos de Peter L. Silveston, de la Universidad de British Columbia, y de E. Lothar Koschmieder, de la Universidad de Texas en Austin, señalan un valor crítico de 1700 ± 50, que es una buena verificación del valor teórico.

El equilibrio entre el empuje de la flotación y las fuerzas disipativas proporciona un criterio para el comienzo de la inestabilidad convectiva, pero ¿qué tipo de movimiento se observará en el fluido cuando la convección esté desarrollada? La teoría de Lord Rayleigh sólo proporciona una limitada respuesta a esta pregunta e incluso teorías más generales, que discutiremos más adelante, no describen todas las propiedades que se observan en la circulación convectiva. Sin embargo, aunque el desarrollo de la convección no puede deducirse cuantitativamente de manera exacta y rigurosa, sí cabe una descripción cualitativa.

En una capa de fluido uniformemente calentada por abajo, el gradiente de temperatura es independiente de la posición horizontal, al igual que la fuerza de flotación resultante. Cuando se supera el valor crítico del número de Rayleigh, el equilibrio inicial pasa a ser inestable; por tanto, el fluido caliente tiende a elevarse por todas partes y el fluido frío tiende a descender igualmente por doquier. Obviamente, ambas cosas no pueden ocurrir a la vez: en un punto el fluido puede ascender o descender, pero no puede moverse en ambos sentidos en el mismo lugar y en el mismo instante. Esta controversia se soluciona con la división espontánea de la capa y el empaquetamiento del fluido en celdas convectivas, en cada una de las cuales el fluido circula en órbitas cerradas.

Existen argumentos teóricos que dan una indicación de la escala más favorable para la formación de estas celdas convectivas. Estos argumentos se fundan en la distinta sensibilidad del estado de estabilidad marginal frente a perturbaciones de diferentes escalas de longitud. La amplitud y la escala de la perturbación son dos magnitudes que no deben confundirse, la primera corresponde al desplazamiento vertical de una región de fluido y la segunda mide el tamaño de dicha región. Para que la teoría de Rayleigh conduzca a resultados satisfactorios, la amplitud debe ser siempre infinitesimal, mientras que la escala puede ser tan grande como permitan las dimensiones del recipiente.

Es costumbre expresar la escala de una perturbación en términos de un número de onda que es el inverso de una longitud. Esta costumbre se debe a que la geometría de una perturbación es generalmente complicada, de modo que no tiene un tamaño claramente definido; sin embargo, la perturbación puede descomponerse en un espectro de modos fundamentales o frecuencias espaciales, al igual que un complejo sonido puede analizarse como una combinación de tonos puros. Un número de onda representa la contribución de una escala de longitud particular en las fluctuaciones internas. Mayores números de onda corresponden a menores escalas o sea a celdas de anchura menor.

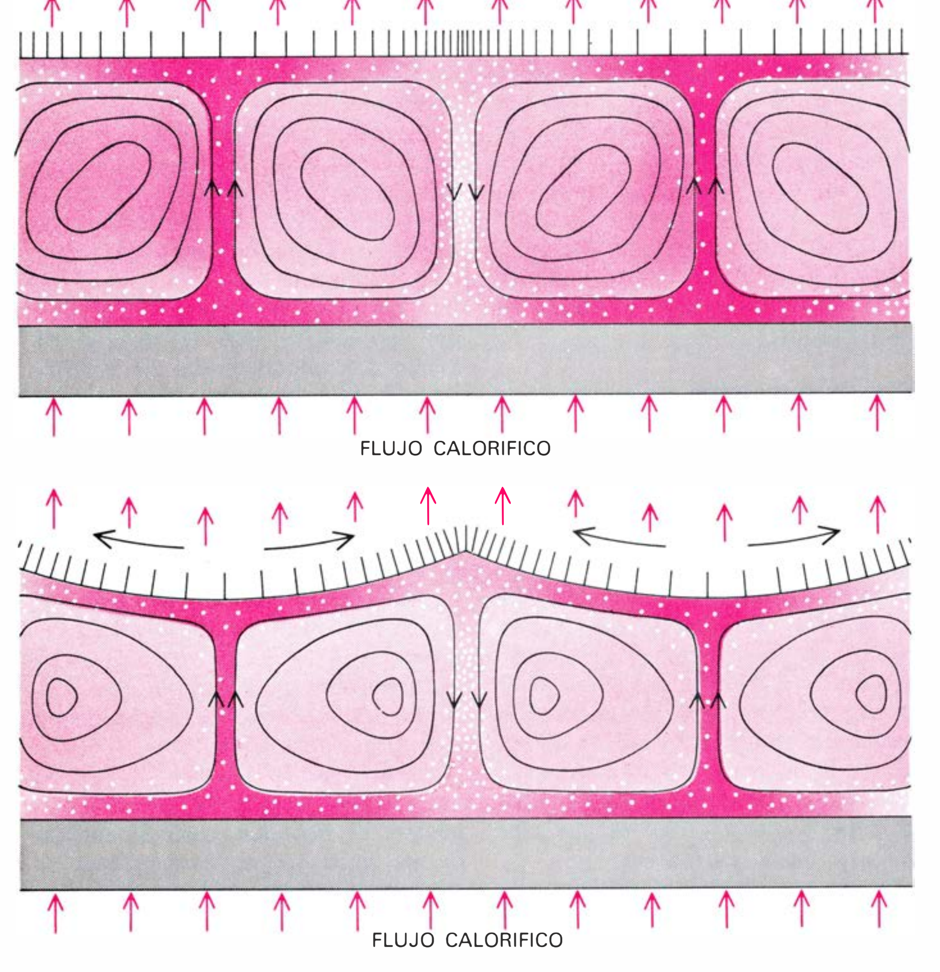
La estabilidad del estado estacionario de reposo es más vulnerable frente a perturbaciones de ciertos números de onda que frente a otros. Se puede imaginar un experimento en el que el número de Rayleigh crítico se midiese en un fluido cuyas fluctuaciones se pudieran controlar, de forma que estuviesen caracterizadas por un único número de onda. Tal experimento revelaría que la inestabilidad comienza antes o más de prisa para el número de onda que describe las perturbaciones con una dimensión horizontal aproximadamente dos veces el espesor de la capa. Para números de onda mayores o menores se necesitan excitaciones más

fuertes (un mayor número de Rayleigh) para provocar la convección. El valor de 1708 para el número de Rayleigh crítico corresponde a las fluctuaciones de tamaño óptimo en la estabilización del estado de reposo.

La sensibilidad del fluido frente a perturbaciones de una escala determinada implica que estas perturbaciones se amplificarán más rápidamente que cualesquiera otras cuando el estado de reposo comience a ser inestable. Por tanto, al iniciarse la convección, es de esperar que la corriente tenga las características de esta escala. No es evidente que estas características persistan una vez que el flujo convectivo se encuentre totalmente desarrollado, aunque esto ocurre siempre que el número de Rayleigh no sobrepase excesivamente el valor crítico.

El número de onda especifica la escala general del flujo, pero no da detalles de su forma; pueden construirse celdas convectivas geométricamente diferentes correspondientes a un mismo número de onda. La estructura convectiva depende de la geometría del recipiente y todavía no ha sido deducida a partir de principios fundamentales, aunque existen reglas empíricas que facilitan predicciones cualitativas.

En el experimento ideal que estamos considerando, en el cual las superficies superior e inferior están delimitadas por contornos rígidos, el elemento básico en la estructura convectiva es un "rollo" en forma tubular alargada. El fluido caliente asciende por uno de los laterales del rollo, alcanza la superficie superior perdiendo calor mientras se desplaza por esta superficie y desciende, a continuación, hacia el fondo de la capa por el lateral opuesto al que había ascendido. La circulación conduce el fluido por el contorno inferior donde su temperatura aumenta de nuevo y recomienza el ciclo. Los rollos adyacentes tienen sentidos de rotación opuestos.



VARIACION DE LA TENSION SUPERFICIAL a lo largo de una interfase. Dicha variación altera la forma de la superficie y de la estructura de la corriente convectiva. La tensión superficial aumenta cuando la temperatura disminuye. Por tanto, una diferencia de temperatura entre puntos de la superficie libre origina un gradiente de tensión superficial. El estado de reposo es inestable si las fuerzas debidas a dicho gradiente superan los efectos disipativos de la viscosidad y la difusión calorífica molecular. En la figura, la magnitud de la tensión superficial se representa por la densidad de empaquetamiento de las barras verticales. El fluido es arrastrado sobre la superficie hacia las regiones más frías, de mayor tensión superficial, siendo reemplazado por fluido caliente que asciende desde estratos inferiores.

La sección transversal del rollo es cuadrada: la base y la altura son iguales y, de hecho, la altura está determinada por el espesor de la capa de fluido. Es decir, las proporciones del rollo son constantes, aunque su lado depende del espesor de la capa. Dado que la unidad básica que se repite no es un rollo, sino dos rollos que rotan en sentidos opuestos, la longitud característica de la celda convectiva es dos veces el espesor de la capa. Esta escala de longitud está en concordancia con las predicciones del análisis teórico.

La planta del modelo celular (su estructura vista desde arriba) viene determinada por la geometría del recipiente donde se realice el experimento. Dado que es muy difícil introducir los detalles de esta geometría directamente en la teoría, ésta no proporciona información alguna sobre la planta del modelo celular; sin embargo, las observaciones experimentales dan una amplia información sobre lo que cabe esperar de esa planta. Así, si el recipiente tiene base rectangular, los rollos se alinean paralelos al lado más corto del rectángulo. En un recipiente de base circular los rollos forman roscos concéntricos. La anchura de cada rollo, y por tanto el número de éstos en el experimento, están determinados por la altura de la capa.

Como indicamos anteriormente, el análisis de convección debido a Lord Rayleigh intentaba explicar las observaciones realizadas por Bénard en torno al año 1900. Actualmente, se sabe que la teoría de Rayleigh no es la apropiada para la convección estudiada por Bénard. Las condiciones experimentales de Bénard difieren, en un pequeño pero sutil detalle, de las del experimento ideal descrito más arriba, y esta diferencia provoca un cambio radical en la estructura convectiva. En la convección de Bénard, los rollos, impuestos por la geometría del problema, pueden aparecer fugazmente cuando comienza la convección, pero posteriormente estos rollos desaparecen dando paso a una geometría más complicada: un mosaico poligonal sobre la superficie del fluido. Al comienzo de la formación de este mosaico los polígonos son irregulares, con un número de lados generalmente comprendido entre cuatro y siete, aunque la mayoría son hexágonos. Cuando la corriente está totalmente desarrollada se alcanza una estructura casi perfecta de hexágonos regulares empaquetados como un panal de colmena. El fluido caliente asciende por el centro de la celda, se extiende por la superficie superior y

cae por el perímetro del hexágono junto a las células adyacentes.

En los experimentos de Bénard, al igual que en el experimento ideal descrito antes, el sistema está constituido por una delgada capa de fluido calentada por la superficie inferior. Pero mientras que en este último el fluido está confinado entre dos placas rígidas, en los experimentos de Bénard la placa superior no existe, de forma que el fluido permanece al aire libre. Puesto que la superficie superior es libre, la tensión interfacial fluido-aire puede afectar a la corriente; de hecho en la convección de Bénard la tensión interfacial es el mecanismo dominante, siendo su variación superficial más importante aún que el empuje de flotación de Arquímedes. Por tanto, no debe sorprendernos que la teoría de Rayleigh, en la cual se supone explícitamente que la única fuerza operante es la flotación, sea inadecuada para explicar la convección de Bénard. Las predicciones de la teoría de Rayleigh en este caso no sirven ni para predecir el valor del gradiente de temperatura necesario para el inicio de la corriente convectiva. No se entendió el problema hasta 1958 en que J. R. A. Pearson, de la Universidad Imperial de Ciencia y Tecnología de Londres, introdujo una teoría alternativa que explicó la convección de Bénard.

La tensión superficial es la fuerza de cohesión entre las moléculas del fluido en la superficie, cuyo efecto neto es minimizar el área de la superficie del fluido. Por ejemplo, en ausencia de otras fuerzas, una gota de líquido tiende a adoptar una forma esférica, puesto que ésta es la configuración de área mínima. La tensión superficial puede imaginarse como una red de tiras elásticas que se extienden en todas direcciones sobre la superficie libre. Si, en algún punto de la superficie, las fuerzas ejercidas por las distintas tiras elásticas que convergen en él no están en equilibrio, la capa superficial se moverá hacia la región de mayor tensión superficial. Este movimiento convectivo superficial se comunica al interior del volumen debido justamente al arrastre viscoso que tiende a anular el movimiento relativo entre distintas capas de fluido.

La tensión superficial, o más precisamente su variación, puede actuar como fuerza impulsora de la convección, ya que, al igual que la densidad, disminuye cuando la temperatura aumenta. Por tanto, un gradiente de temperatura a lo largo de la superficie del fluido induce un gradiente de tensión superficial. Las regiones más frías tendrán mayor tensión superficial, mientras que en las más calientes ésta será menor. Si el gradiente de tensión superficial provoca un desequilibrio de fuerzas en la superficie, éste originará el movimiento convectivo.

El comienzo de la inestabilidad convectiva en el experimento de Bénard puede analizarse en forma análoga a la del caso inducido por empuje de Arquímedes. Supongamos un elemento de fluido caliente que se desplaza verticalmente hacia arriba debido a una pequeña fluctuación interna. Independientemente de la existencia del empuje de flotación y de que éste refuerce o disminuya el movimiento, se producirá en la superficie un pequeño aumento de la temperatura, justo por encima del lugar donde se formó la fluctuación; esta variación de la temperatura generará una disminución de la tensión superficial en este punto. Aún así, las fuerzas en la superficie continuarán en equilibrio, ya que la superficie de los alrededores de este punto "tira" por igual de este punto en todas las direcciones. Para que el flujo se inicie, es necesaria una segunda perturbación que cause un desplazamiento horizontal del elemento fluido de la superficie en la zona de tensión superficial reducida. En este caso, las fuerzas de tensión superficial que actúan sobre este pequeño elemento no están en equilibrio y, si el gradiente de tensión superficial es lo suficientemente grande, el desplazamiento se verá amplificado. El elemento de superficie se verá arrastrado, a lo largo de la superficie, hacia la zona más fría de mayor tensión superficial, arrastrando consigo por viscosidad parte del volumen del fluido. Aparece así un hueco en la región donde se encontraba inicialmente que viene a ser ocupado por el fluido más caliente que se halla en los estratos inferiores, de modo que se verán reforzados los gradientes de temperatura y de tensión superficial. Mientras tanto, el elemento superficial de fluido se enfría en su recorrido a lo largo de la superficie y empieza a caer, estableciéndose la corriente convectiva.

Al igual que en la convección inducida por flotación, la mera existencia de un gradiente de temperatura no garantiza que vaya a desarrollarse un flujo convectivo. El gradiente debe ser suficientemente grande como para vencer los efectos disipativos de viscosidad y difusión de calor. La relación entre estos mecanismos se expresa mediante otra magnitud adimensional, en este caso denominada número de Marangoni, en honor del científico italiano del siglo XIX C. G. M. Marangoni. La expresión del número de Marangoni es la misma que la del número de Rayleigh, excepto en que la fuerza de flotación se sustituye por la fuerza inducida por tensión superficial; en otras palabras, el número de Marangoni es el cociente entre el gradiente de tensión superficial y el producto del frenado viscoso por la difusión calorífica en el medio fluido. La convección de Bénard comienza cuando el número de Marangoni supera un cierto valor crítico.

Una propiedad intrínseca de la convección inducida por los gradientes de tensión superficial es que modifica el contorno de la superficie. Las zonas que tienen mayor tensión superficial tienden a encogerse, reduciendo su área total, algo que desafía nuestras nociones intuitivas. En el centro de una celda de Bénard, por donde el fluido sube, la superficie presenta una depresión y en los bordes de la célula, por donde el fluido cae, presenta una elevación. Las fuerzas gravitatorias se oponen a la formación de estas ondulaciones en la superficie libre, puesto que la energía potencial gravitatoria para las pequeñas dimensiones de los experimentos de Bénard es mínima para el caso en que la superficie superior es plana. Vemos, pues, que las interacciones entre gravitación y tensión superficial son sutiles y complejas. Fue en 1964 cuando D. A. Nield, de la Universidad de Auckland, incorporó ambas fuerzas a la teoría de la convección de Bénard.

Una demostración inequívoca de la supremacía de la tensión superficial en la convección de Bénard es que la forma hexagonal de las celdas convectivas aparece incluso cuando la capa de fluido se calienta por arriba en vez de por abajo. Pero cuando se calienta por arriba, el gradiente de densidad se opone al movimiento convectivo, de manera que las fuerzas inducidas por gradientes de tensión superficial deben vencer, no sólo el frenado viscoso y la difusión de calor, sino también dicho gradiente de densidad. De igual manera, se han observado flujos convectivos atribuidos a gradientes de tensión superficial en los experimentos realizados durante dos misiones espaciales Apolo, en situaciones en que la gravedad y, por tanto, las fuerzas de flotación son irrelevantes.

La teoría de Rayleigh, así como las diferentes variantes de ésta, dan aproximaciones sobre las condiciones suficientes para el inicio de la convección. Pero, ¿qué ocurre una vez esta-

blecido el flujo convectivo? Estas teorías son poco satisfactorias al describir la convección totalmente desarrollada.

En la teoría de Rayleigh, la velocidad de la corriente viene expresada como una función exponencial: la velocidad es proporcional al número de Euler, *e* (cuyo valor númerico es aproximadamente 2,7), elevado a cierta potencia. El exponente es igual al tiempo (medido en segundos a partir de un instante de referencia) multiplicado por un coeficiente, λ (que mide la "distancia" al número de Rayleigh crítico). Es decir, la velocidad es proporcional a $e^{\lambda t}$. Las predicciones de la teoría pueden obtenerse observando la evolución con el tiempo de la expresión $e^{\lambda t}$ para diferentes valores de λ.

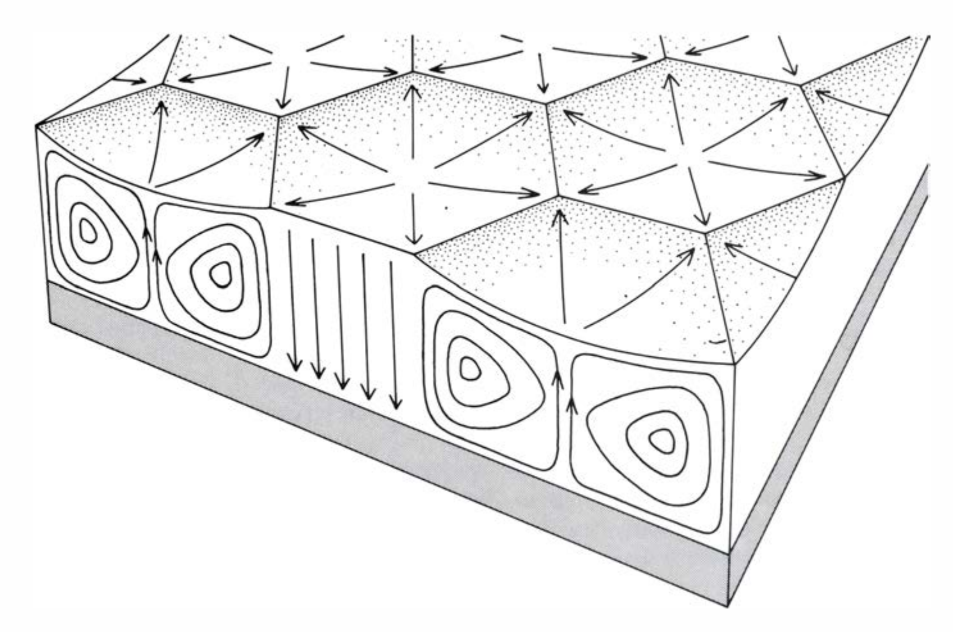
Si el número de Rayleigh es menor que el valor crítico, λ será negativo. Por tanto, en el curso del tiempo, la velocidad viene dada por potencias negativas de *e*, cuyo exponente aumenta, en valor absoluto, progresivamente con el tiempo; bajo estas condiciones, el valor de la exponencial se aproxima a cero según aumenta *t*. Es decir, la velocidad tiende a cero y cualquier movimiento interno del fluido será anulado por los mecanismos disipativos. Cuando el número de Rayleigh es exactamente igual a su valor crítico, λ será cero, y así el exponente λt será nulo (cero) en cualquier instante; puesto que cualquier número elevado a cero es igual a 1, la perturbación ni se amortigua ni se amplifica, sino que mantiene siempre el mismo valor inicial.

Estas dos predicciones concuerdan con la intuición y con el análisis de estabilidad desarrollado anteriormente. Un valor negativo de λ corresponde a un estado de reposo estable y el valor $\lambda = 0$ designa el estado de estabilidad marginal. Sin embargo, cuando el valor del número de Rayleigh sobrepasa el valor crítico, es decir, cuando puede desarrollarse una corriente convectiva, la teoría es inadecuada. En efecto, en este caso λ será positivo y la potencia a que está elevado *e* aumentará continuamente con el tiempo, de forma que la velocidad tenderá a crecer exponencialmente sin límite. Por ejemplo, si λ es igual a + 1 y la velocidad inicial igual a un centímetro por segundo, al cabo de un segundo la velocidad habrá aumentado a 2,7 centímetros por segundo y después de dos segundos la velocidad será de 7,4 centímetros por segundo. La velocidad crece sin límites, lo que conduce a lo imposible: por ejemplo, la corriente convectiva sobrepasaría la velocidad de la luz en menos de medio minuto.

La dependencia de la velocidad de la corriente convectiva respecto del valor de λ se ha presentado aquí en una forma simplificada. En general, λ es un número complejo, que tiene una parte real y una parte imaginaria; incluyendo esta última la raíz cuadrada de menos uno como factor multiplicativo. Sólo hemos considerado la variación de la parte real de λ; pero si la parte real de λ es igual a cero y la parte imaginaria diferente de cero (condición denominada de sobrestabilidad) pueden aparecer movimientos convectivos oscilatorios (periódicos en el tiempo). Tales oscilaciones han sido observadas en el laboratorio y constituyen una interesante subclase de fenómenos convectivos. Sin embargo, en el modelo correspondiente a la teoría de Rayleigh, la parte imaginaria de λ es siempre nula y, por tanto, debe afrontarse el problema del crecimiento exponencial.

Evidentemente, la velocidad de la convección natural no puede seguir un crecimiento exponencial. Por este motivo, las predicciones de la teoría de Rayleigh pueden considerarse válidas sólo cuando, o bien el número de Rayleigh es muy próximo al valor crítico (de modo que λ sea pequeño), o bien para un corto intervalo de tiempo después del inicio de la convección (de forma que *t* sea pequeño). Estas limitaciones son debidas a las hipótesis que se consideran en la teoría. En particular, se supone que el gradiente de temperatura es constante y, por tanto, no se ve afectado por la circulación convectiva. Esta suposición es claramente contradictoria con la realidad; pues cuando el fluido caliente asciende y se introduce en la zona superior más fría, se reduce la diferencia de temperaturas entre la placa superior e inferior; de manera que el empuje de flotación disminuye en la misma proporción en que se reduce la diferencia de temperaturas. La corriente, en consecuencia, se autolimita. En cambio, en la teoría de Rayleigh no aparece tal autolimitación. En ésta, independientemente de la velocidad de la corriente convectiva, el gradiente de temperatura permanece constante; por tanto, el empuje de flotación se mantiene constante y comunica a la corriente una aceleración también constante que causa el aumento continuo e ilimitado de la velocidad.

Dada esta contradicción, puede parecer sorprendente que la teoría de Rayleigh conduzca a resultados aceptables bajo determinadas circunstancias; esto se debe a la hipótesis de que la región de fluido sólo puede realizar un desplazamiento infinitesimal, por lo que considerar que el gradiente de temperatura no varía, es bastante razonable. Aun cuando la región de fluido se desplace una distancia finita y pequeña causará, únicamente, una pequeña perturbación en la distribución de temperaturas, de modo que siguen siendo aproximadamente válidas las predicciones de la teoría. Estas suposiciones no se verifican en el caso de una corriente convectiva totalmente desarrollada; por este motivo, la



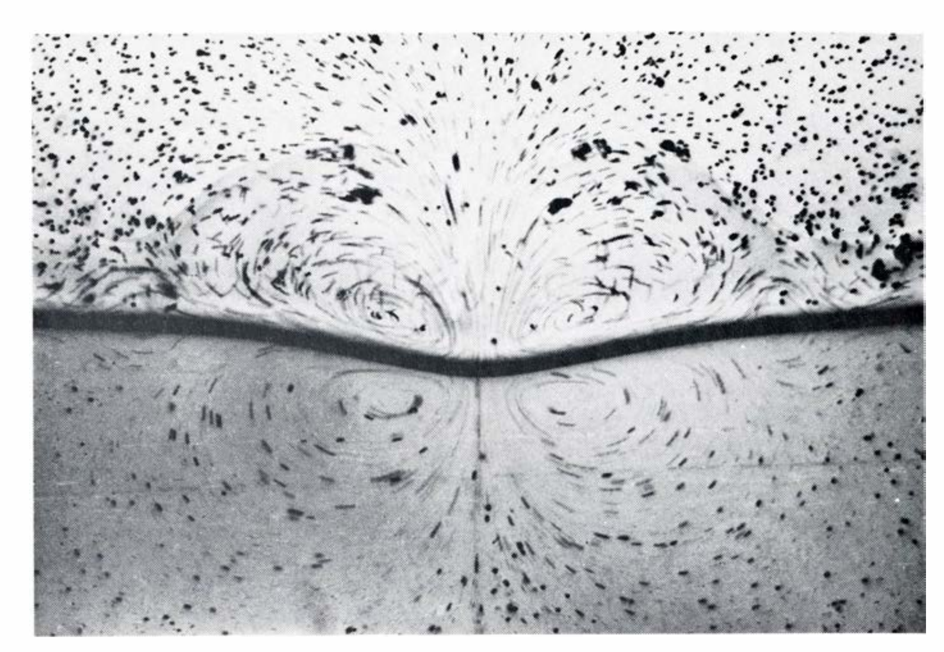
LAS CELDAS HEXAGONALES de la superficie libre del fluido forman un mosaico característico de la convección provocada por gradientes de tensión superficial. En los puntos donde la tensión es mayor, la superficie se encoge, reduciendo así su área parcial. En el centro de las celdas por donde el líquido asciende (desciende en el caso de gases), la superficie tiene una depresión (elevación en los gases); de modo que el líquido debe superar una pequeña pendiente sobre la superficie, para descender, a continuación, por el perímetro de la celda hacia la placa inferior. (Los dibujos de este artículo son de Alan D. Iselin.)

teoría de Rayleigh predice aquí una velocidad infinita de la corriente, y esto es imposible.

Una teoría válida para la convección debe de algún modo incluir esta autolimitación por la cual la corriente misma altera la fuerza que la provoca. No existe ningún método para resolver el problema rigurosa y exactamente; hay, por supuesto, aproximaciones más completas que la teoría de Rayleigh. La teoría que discutiremos a continuación se basa en las ideas introducidas en 1937 por el físico ruso L. D. Landau para describir ciertos tipos de transiciones de fase, tales como el inicio de la imanación en el hierro. Landau, junto a otro físico ruso, V. L. Ginzburg, adaptaron, años más tarde, esta teoría para la descripción de la superconductividad en los metales. Ambos fenómenos tienen una característica en común con la convección: requieren la descripción simultánea de fluctuaciones de diferentes escalas de longitud en el sistema. Cuando se adapta la teoría de Landau al problema de la convección, ésta incluye a la teoría de Rayleigh como una primera aproximación.

La teoría debe proporcionar una ecuación de movimiento para el fluido, dando la velocidad y aceleración de una región del fluido para cualquier combinación específica de las condiciones que se pueden imponer desde el exterior al sistema. Dicha ecuación de movimiento puede construirse directamente; mas, para entender con mayor claridad el significado de la teoría de Landau, es preferible definir una superficie potencial, a partir de la cual puede obtenerse la ecuación de movimiento. (Esta es simplemente la ecuación que da la pendiente de la superficie potencial en cada punto.)

La superficie potencial puede imaginarse como un paisaje ondulado, donde en un punto determinado la altura por encima o por debajo de cierto plano de referencia representa la energía relativa del estado del fluido correspondiente a dicho punto. La tendencia de la energía de alcanzar un valor mínimo implica que el punto que representa el estado de sistema "rodará cuesta abajo" siempre que sea posible. Un eje sobre el plano de referencia define la línea de velocidad cero, los desplazamientos a derecha o izquierda de esta línea denotan un aumento positivo (hacia arriba) o negativo (hacia abajo) de la velocidad del fluido. La posición en este eje de velocidad cero viene determinada por la diferencia existente entre el número de Rayleigh y el valor crítico de éste; a esa diferencia la designaremos por ΔR. Así, los



UNA REGION CALIENTE que suba y alcance la interfase que separa los líquidos crea en ésta un gradiente de tensión superficial. La depresión sobre la superficie indica el punto en que la tensión superficial es mínima. El líquido es arrastrado hacia derecha e izquierda de dicho punto y el movimiento superficial se comunica al volumen, debido al arrastre viscoso. El mismo mecanismo opera en un fluido cuya superficie superior es libre y abierta al aire. La fotografía es de H. Linde, del Instituto Central de Físico-Química de Berlín (RDA). Para visualizar la corriente se suspendieron partículas de plástico.

valores de ΔR y de la velocidad V, especifican un punto sobre el plano de referencia (V, ΔR) y la altura de la superficie potencial en este punto indica la energía convectiva del sistema en ese estado (determinado por un gradiente de temperatura dado por ΔR y una velocidad V).

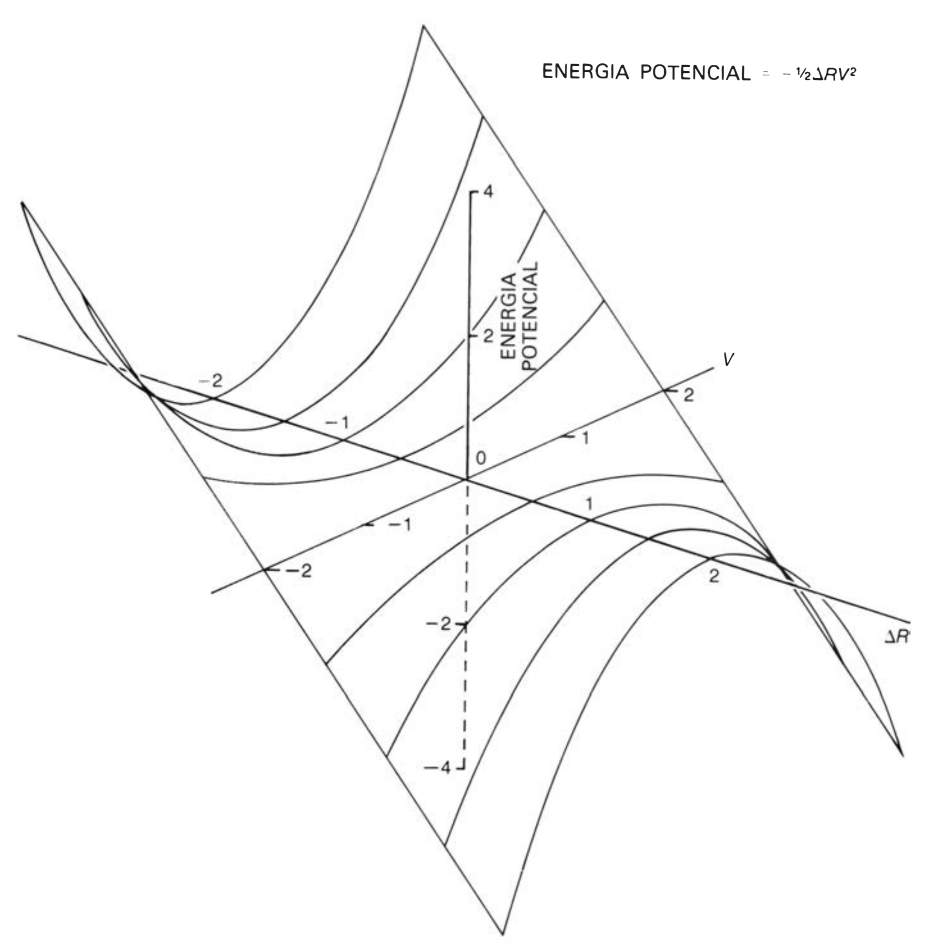
Conviene discutir con cierto detalle las propiedades de la ecuación que determina la topografía de la superficie potencial. Dicha ecuación puede escribirse como una suma infinita de términos, incluyendo sucesivamente cada uno de estos términos una potencia superior de la velocidad de corriente. El primer término es cuadrático en V: $-\frac{1}{2}\Delta R V^2$. En el segundo término aparece V^3 multiplicado por un cierto coeficiente, el siguiente sería en V^4, y así sucesivamente. Todas y cada una de estas potencias está multiplicada por un cierto coeficiente que mide su contribución a la forma de la superficie.

Como puede suponerse, no sería posible calcular la suma de los infinitos términos de la serie, aun sabiendo que existen métodos formales para realizar dicha suma, puesto que no todos los coeficientes son conocidos. Sin embargo, es de esperar que los coeficientes sean cada vez menores según aumente la potencia a que V esté elevada. Asimismo, y basándonos en la hipótesis anterior, cabe esperar que se obtenga una válida aproximación truncando la serie y despreciando, por tanto, todos los términos que contienen potencias superiores a una dada. Si la velocidad de la corriente no es demasiado grande es razonable imaginar que sea pequeña la contribución de los términos a partir de dicha potencia. En particular, usando un sistema adecuado de unidades en el cual la velocidad sea menor que la unidad, las potencias superiores de V convergen hacia cero y la aproximación será válida.

La teoría de Rayleigh puede verse como la truncación de la serie infinita en la que se mantiene únicamente el primer término que vale $-\frac{1}{2}\Delta R V^2$. La superficie descrita por esta expresión cuadrática tiene dos lóbulos, cada uno con una sección transversal parabólica; un lóbulo es cóncavo hacia arriba y el otro es convexo (cóncavo hacia abajo). Resulta obvio, a partir de la ecuación, que siempre que V sea cero (es decir, en cualquier punto sobre el eje de velocidad cero), el valor de la energía relativa también será cero. Si el número de Rayleigh es subcrítico, ΔR será negativo y la superficie se elevará hacia ambos lados del eje de velocidad cero, de forma que la energía aumentará siempre que la velocidad sea diferente de cero. Es decir, el estado de reposo corresponde a un estado de equilibrio estable. Cuando el número de Rayleigh es supercrítico, ΔR es positivo, y ocurre la situación opuesta. La superficie desciende a ambos lados del eje de velocidad cero, que denota ahora la línea de energía máxima y, por tanto, los estados de equilibrio inestable.

Estas propiedades de la superficie potencial son las descritas por el análisis elemental de estabilidad presentado anteriormente, e incorporan todas las ventajas e inconvenientes de la teoría de Rayleigh. En un entorno próximo al origen, donde tanto ΔR como V son pequeños, el comportamiento del sistema puede deducirse a partir del cambio en la curvatura de la superficie. Cuando ΔR es negativo, el fluido vuelve al estado de reposo tras sufrir una pequeña perturbación; pero si ΔR es positivo, la perturbación se amplifica y comienza la corriente convectiva. Cuando ΔR es igual a cero (el número de Rayleigh es igual al valor crítico), la superficie potencial es localmente plana y cualquier fluctuación interna de la velocidad ni se amplifica ni decae. Para mayores valores de ΔR y V, nos encontramos de nuevo con el problema de que la velocidad crece indefinidamente, puesto que la superficie potencial "cae" hacia menos infinito.

En la teoría de Landau, este problema puede solucionarse truncando la serie infinita en una potencia superior a dos. De hecho, es posible obtener importantes mejoras con sólo añadir algunos términos más. Consideraciones de simetría en el experimento para el cual se ha desarrollado la teoría de Rayleigh reducen la elección de términos. En dicho experimento la forma de la corriente debe ser la misma tanto si calentamos por abajo como si enfriamos por arriba; o sea, si, para un mismo gradiente de temperatura, todas las velocidades invierten su sentido. Esta invariancia implica que la superficie potencial es simétrica con respecto al eje ΔR (de velocidad cero); tal superficie vendrá descrita por una ecuación en la que se incluyen tan sólo términos en potencias pares de la velocidad (tales como V^2, V^4 y sucesivas). Siempre que el exponente es un número par, V y $-V$ elevadas a la misma potencia dan el mismo resultado, mientras que los resultados tienen signos opuestos cuando el exponente es impar; por tanto, los coeficientes de V^3 y demás potencias impares deben ser nulos.



LA SUPERFICIE POTENCIAL asociada a la teoría de Rayleigh define la energía relativa del fluido en función del número de Rayleigh y de la velocidad del fluido. La forma de la superficie es una ley cuadrática $-1/2\ \Delta RV^2$; siendo el coeficiente ΔR la diferencia entre el número de Rayleigh y el valor crítico de éste. Cuando ΔR es negativo, el eje de velocidad cero (denotado por ΔR) designa el estado de mínima energía y cualquier movimiento del fluido ($V \neq 0$) correspondería a un estado con mayor energía que el estado de reposo, siendo éste, pues, un estado de equilibrio estable. Cuando ΔR es positivo, se invierte la pendiente de la superficie y el eje ΔR (de velocidad cero) define un conjunto de estados de equilibrio inestable; el fluido puede reducir su energía entrando en convección. El mayor defecto de la teoría de Rayleigh consite en que, una vez que la corriente ha comenzado, predice un aumento continuo e ilimitado de la velocidad de dicha corriente, predicción que, a todas luces, resulta irrealizable.

Se obtienen resultados interesantes al incluir un término más, aparte del término cuadrático, en el potencial de Landau. Es decir, incluyendo el término en orden cuatro, que por comodidad tomamos como $-\frac{1}{4}V^4$; de manera que la superficie potencial viene determinada por la expresión $-\frac{1}{2}\Delta RV^2 + \frac{1}{4}V^4$. Cuando ΔR es negativo, la superficie potencial es muy semejante a la del potencial anterior, aunque la energía aumenta más rápidamente cuando la velocidad se aleja de cero. En cambio, cuando ΔR es positivo, la forma de la superficie es totalmente distinta. La energía disminuye a cada lado del eje de velocidad cero, pero este decaimiento no continúa indefinidamente, ya que la energía alcanza un valor mínimo (para una cierta velocidad de la corriente convectiva) y a continuación aumenta si la velocidad sigue creciendo. Tanto la profundidad del mínimo de energía como la velocidad para la que se alcanza dicho mínimo aumentan a medida que crece ΔR.

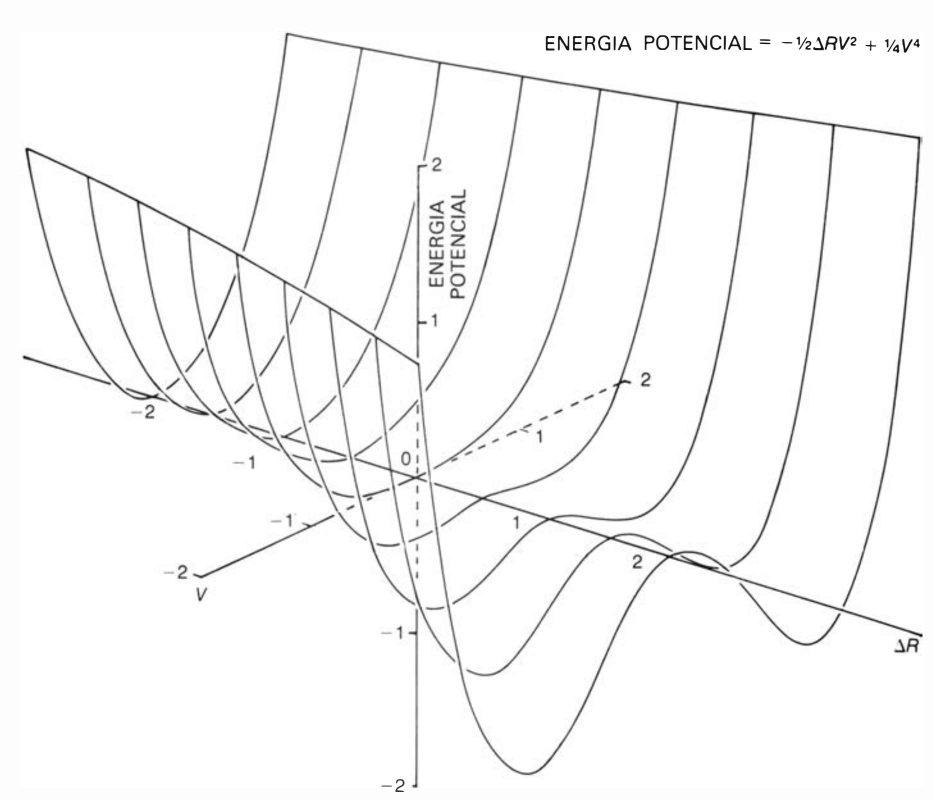
De nuevo esta versión de la teoría de Landau (incluyendo únicamente los términos de segundo y cuarto orden) conduce a predicciones realistas sólo para un rango limitado de valores del número de Rayleigh y de la velocidad de la corriente. Como antes, cuando el número de Rayleigh es menor que el valor crítico cualquier fluctuación interna de la velocidad decae con el tiempo y el estado estacionario de reposo es el estado de energía mínima y de equilibrio estable. Para valores supercríticos del número de Rayleigh, hay perturbaciones que se amplifican, pero no de forma continua y sin límite. Cuando la velocidad alcanza un cierto valor finito, determinado por ΔR, y correspondiente al mínimo de la superficie potencial, se alcanza un nuevo estado de equilibrio estable. Cualquier separación de este estado es amortiguada por los efectos disipativos, y el sistema vuelve a tener al cabo de cierto tiempo el valor de la velocidad correspondiente a este nuevo estado de equilibrio.

La teoría de Landau con un potencial de cuarto orden elimina algunos defectos de la teoría de Rayleigh, pero sigue siendo una aproximación, y, por tanto, sólo válida cuando la velocidad no es demasiado grande. Cuando V es grande, las potencias superiores de V aportan una contribución importante aunque el coeficiente que las precede sea pequeño; por este motivo, una

teoría que desprecia todos los términos superiores al cuarto no puede representar satisfactoriamente la forma de la superficie potencial lejos del eje de velocidad cero, pues, para grandes valores de la velocidad, se presenta la delimitación del dominio de la estabilidad del propio régimen convectivo, problema que con el potencial de orden cuarto no cabe estudiar de una manera realista. Por otra parte, en algunas situaciones existe un sentido de la corriente que está más favorecido que el otro, de forma que se pierde la simetría de la superficie potencial y deben incluirse potencias impares de V en la ecuación. Tal es el caso, por ejemplo, de los fluidos cuyas propiedades (viscosidad, etcétera) dependen de la temperatura.

Tanto la teoría de Landau como la de Rayleigh se derivan de experimentos ideales en los que las propiedades del fluido son constantes (excepto la temperatura y la densidad). Los fluidos reales son más complicados y las relaciones entre las distintas propiedades del fluido pueden ser bastante complejas. Por ejemplo, en el modelo se supuso que únicamente la densidad variaba con la temperatura. En realidad la viscosidad y la difusividad térmica también cambian con la temperatura en los fluidos reales. Puesto que dichas cantidades entran en la definición del número de Rayleigh, cualquier variación de éstas puede tener un influjo decisivo en el inicio de la convección y su evolución posterior. Se supuso, también, que el fluido era incompresible, y como sabemos que muchos fluidos son compresibles, la presión ha de jugar en ocasiones un papel importante; de hecho, la presión influye sobre la densidad y otras propiedades del fluido. La relación entre la viscosidad y la temperatura es bastante compleja. En general, la viscosidad disminuye al aumentar la temperatura, pero al mismo tiempo la energía disipada por las fuerzas viscosas fluye como calor, de manera que aumenta la temperatura del sistema.

Una teoría que tuviese en cuenta explícitamente todas las relaciones conocidas entre las propiedades del fluido sería demasiado compleja, difícil de manejar y poco práctica. A veces, para describir un fluido real es más importante buscar una relación óptima entre la complejidad del problema y de la teoría, que intentar buscar un modelo matemático riguroso, pero excesivamente complicado. La necesidad de este compromiso puede ilustrarse con algunos ejemplos de convección fuera del laboratorio.



LA TEORIA DE LANDAU incluye términos adicionales en la ecuación que define la superficie potencial, haciéndola más realista. Aquí se muestra el caso más sencillo en que se añade un término de cuarto orden al término cuadrático de la teoría de Rayleigh. Al igual que en ésta, el estado estacionario de reposo es inestable cuando ΔR es mayor que cero, aunque ahora la velocidad no continúa aumentando indefinidamente, sino que tiende hacia un valor finito correspondiente a un mínimo en la superficie potencial. La teoría predice, pues, un nuevo estado de equilibrio estable con una velocidad de corriente tal que minimiza la energía total. Esta teoría fue ideada en 1937 por L. D. Landau para describir las transiciones de fase en materiales magnéticos y ha sido recientemente adaptada a la convección.

En la atmósfera terrestre, la convección se observa a muy diversas escalas. Por ejemplo, el gradiente de temperatura entre los trópicos y los polos da origen a una circulación global que puede descomponerse en tres grandes celdas convectivas por hemisferio. Estas celdas, modificadas por la rotación de la Tierra, parecen ser causa de los vientos alisios en los trópicos y de los correspondientes contralisios. El calentamiento local de la superficie terrestre origina corrientes convectivas a menor escala, estando incluidas la mayoría de las tormentas dentro de este grupo de corrientes convectivas locales. Los cúmulos, nubes que se forman cuando asciende el aire húmedo caliente y se enfría pasando a ser sobresaturado, condensándose el vapor de agua que contienen, señalan, a menudo, la inestabilidad atmosférica.

Un análisis teórico de la convección en la atmósfera debe tener en cuenta la gran comprensibilidad del aire, que provoca un gradiente de densidad, incluso cuando la temperatura es constante con la altura. Una descripción adecuada de la convección atmosférica debe incluir, también, el calentamiento adiabático del aire cuando se introduce en una zona a mayor presión. La viscosidad y otras propiedades del aire varían con la temperatura y con la presión; la presencia de vapor de agua, que produce calor cuando se condensa, aumenta el nivel de complejidad del problema. Las nubes formadas como resultado de una corriente convectiva son, frecuentemente en sí mismas, inestables frente a posteriores movimientos convectivos: la nube se enfría por la parte superior, perdiendo energía por radiación y se calienta por la parte inferior, a través de la radiación absorbida desde la tierra. Si la magnitud de estos efectos es suficientemente grande, puede llegar a desarrollarse una celda convectiva en el interior de la propia nube.

A pesar de todas estas complicaciones, el movimiento convectivo en la atmósfera responde, generalmente, al mismo modelo básico utilizado para describir sencillos experimentos de laboratorio. Las formaciones lineales de nubes, paralelas unas a otras como los rollos convectivos de Lord Rayleigh, es un hecho corriente que puede observarse cuando se vuela en avión; las fotografías desde satélites artificiales revelan formaciones de celdas poligonales que se extienden sobre miles de

kilómetros cuadrados. Sin embargo, en este caso, los resultados del experimento en el laboratorio no pueden extrapolarse a escala atmosférica. Las celdas convectivas desarrolladas en el laboratorio tienen siempre la misma altura que anchura, mientras que las celdas convectivas atmosféricas son mucho más anchas, aproximadamente en un factor 50. Además, el sentido de la circulación en los experimentos de laboratorio a pequeña escala es siempre el mismo (en gases es hacia abajo en el centro de la celda); en la atmósfera, sin embargo, se observan ambos sentidos de corriente.

La convección en los océanos abarca también un amplio rango de escalas, desde un metro o dos hasta el tamaño de las mismas cuencas oceánicas. El más simple de estos movimientos tiene una explicación sencilla. La radiación solar calienta el agua a profundidad considerable, puesto que algunas longitudes de onda penetran decenas de metros en el océano. Por otra parte, el enfriamiento tiene lugar, casi totalmente, por evaporación y por pérdida de calor por conducción y radiación a la atmósfera; ambos procesos están ubicados, principalmente, en la superficie. De esta forma, en los océanos se introduce energía a un nivel inferior al que se pierde, y se puede tener una capa de algunos metros de espesor que sea inestable por causa de esta diferencia de temperatura a lo largo de la vertical.

La compresibilidad del agua del mar es pequeña y puede tener influencia sobre el movimiento convectivo sólo en las grandes profundidades marinas; sin embargo, la salinidad del agua tiene una influencia decisiva. La densidad del agua del mar no sólo varía con la temperatura, también depende de la salinidad, ya que la densidad aumenta al aumentar la cantidad de sal; la interacción entre estos factores (temperatura y salinidad) puede originar nuevos tipos de movimiento convectivo que no se observan cuando está presente un único gradiente, bien sea de temperatura o de salinidad.

Si la temperatura es mayor en la parte inferior de la capa y la salinidad en la parte superior, ambos gradientes se refuerzan entre sí para provocar la convección. Cuando los gradientes de temperatura y salinidad ocasionan efectos opuestos sobre el gradiente de densidad, entran en juego efectos más sutiles. Si agua caliente y salada se encuentra por encima de agua fría y dulce, el gradiente de temperatura favorece la estabilidad, mientras que el gradiente de salinidad induce la inestabilidad. Incluso cuando los dos gradientes se combinan para dar una densidad uniforme, es posible que la convección se desarrolle por motivos relacionados con el modo de disipación de los dos gradientes. El gradiente de temperatura se disipa principalmente por difusión calorífica, mientras que el de salinidad se disipa por difusión másica de las moléculas de sal y de agua. La difusión calorífica es mucho más rápida, aproximadamente unas 100 veces más rápida que la difusión másica. Inicialmente, pueden ajustarse la temperatura y la salinidad de cada capa para que la densidad sea la misma en ambas. Si una región de la capa de fluido caliente y salada desciende y se introduce en la capa de fluido fría y dulce, "perderá" temperatura por difusión calorífica alcanzando el equilibrio térmico con su entorno mucho antes de que la difusión másica haya podido reducir significativamente su salinidad; será, pues, más densa que su entorno y su movimiento se verá amplificado.

La combinación opuesta, con el líquido frío y dulce por encima del fluido caliente y salado, puede conducir al fenómeno de sobrestabilidad convectiva. Una región de agua salada y caliente, al elevarse, se enfría, pero mantiene su concentración de sal, siendo más densa que lo era inicialmente, cayendo así hacia la capa inferior. De hecho, puede sobrepasar su posición inicial y continuar oscilando alrededor de ésta. Estas oscilaciones pueden amortiguarse o amplificarse, dependiendo de la magnitud de los dos gradientes.

Uno de los sistemas convectivos más complejos es el que en el manto terrestre parece estar detrás de la formación de cadenas montañosas en el fondo del Atlántico y de la deriva de los continentes a lo largo de la superficie del globo terráqueo. El calor que da origen a la circulación convectiva no se libera en un contorno, sino más bien en todo el volumen del material, principalmente debido a la desintegración de los elementos radiactivos. Bajo estas circunstancias se origina un gradiente de temperatura, puesto que el sistema pierde calor tan solo por la superficie, de modo que la temperatura aumenta con la profundidad. Hay pocas dudas de que el gradiente sea lo suficientemente grande como para inducir la convección, pero las propiedades del sistema son tan complicadas y el manto terrestre es tan inaccesible a las mediciones directas que la forma y dimensiones del sistema convectivo resultan casi desconocidas. La viscosidad aumenta bruscamente con la profundidad, y a ciertos niveles, en la zona convectiva, el material sufre una transición de una fase cristalina a otra.

A una escala mucho menor puede observarse un interesante proceso convectivo durante el secado de una película de pintura o de laca. En este caso, la fuerza que genera la convección no es el empuje de flotación, sino la tensión superficial, como mencionamos en los experimentos de Bénard. El mecanismo responsable de la corriente convectiva es la evaporación del disolvente desde la superficie libre de la película. Si una perturbación aumenta la velocidad de evaporación en alguna región, ésta se enfriará, incrementando su tensión superficial. Además, la tensión superficial de los pigmentos y demás macromoléculas de la película es, generalmente, mayor que la tensión superficial del disolvente, de forma que una disminución de la cantidad de disolvente en la superficie aumenta la tensión superficial, independientemente de la temperatura. El líquido es arrastrado por la superficie hacia las regiones de mayor tensión superficial; al llegar a éstas, desciende hacia la base de la película y cierra el ciclo. Sin embargo, cuando la concentración del disolvente disminuye, la viscosidad aumenta y, finalmente, el número de Marangoni se hace menor que el valor crítico, deteniéndose la convección.

A menudo, cuando una pintura se seca, las celdas convectivas tienen una forma hexagonal, o al menos una forma poligonal que se aproxima a la forma ideal de hexágonos regulares. La convección puede causar distribuciones anómalas de los pigmentos, dando origen a las irregularidades en el color que se observan en la película seca. En algunos casos, el modelo tridimensional de las celdas convectivas permanece como si estuviese congelado en la película seca. Tal fenómeno no siempre es indeseable: el acabado granulado o gratinado de algunas pinturas adquiere su textura por este mecanismo.

Así pues, hemos visto que los fenómenos convectivos son muy diversos: la circulación atmosférica, las corrientes oceánicas, la circulación celular en pinturas, etcétera. Las teorías que describen estos movimientos en fluidos requieren muchas hipótesis simplificadoras si queremos que sean de utilidad práctica y, aún así, dichas teorías no son todo lo sencillas que uno quisiera. Cuán notable es, sin embargo, en estas teorías que, haciendo uso de unos pocos parámetros adimensionales, quepa el entendimiento y explicación de fenómenos de escalas muy dispares.

Fisiología del koala

Este dócil marsupial come principalmente hojas de eucalipto (tóxicas para otros animales), bebe rara vez y no utiliza refugio. Estudios recientes indican cómo se las arregla para sobrevivir en su nicho característico

Robert Degabriele

Una reacción habitual en las personas que ven un koala (o, más probable, la fotografía de uno de ellos, puesto que el único lugar fuera de Australia donde los koalas viven es en el zoológico de San Diego) es considerar al animal como un osito de peluche vivo y pensar si podrían tener uno en casa. El koala, desde luego, no es ningún oso, ni lo podemos tener en casa. Se trata de un marsupial, un mamífero primitivo cuyas crías nacen en un estado prácticamente embrionario y pasan la primera parte de su vida en una bolsa o marsupio que cubre los pezones de la madre. Y aunque antaño se permitía mantener koalas como animales domésticos, esta costumbre está ahora prohibida por rigurosas leyes de protección que se promulgaron en Australia para defenderlo de los cazadores y traficantes de pieles.

El nombre científico del koala es *Phascolarctos cinereus,* términos que derivan, el primero, del griego, y significa oso con bolsa, y, el segundo, del latín e indica gris ceniza. A este animal se le caracteriza formalmente como un folívoro arborícola, lo que significa que vive en árboles y come hojas. El árbol del koala es el eucalipto, concretamente 35 o más de las aproximadamente 600 especies del género *Eucalyptus* que crecen en Australia. La dieta de un koala adulto está compuesta casi exclusivamente por hojas de eucaliptos, y, puesto que los aceites de estas hojas son tóxicos para la mayoría de los demás mamíferos, uno se pregunta qué adaptaciones permiten al koala medrar gracias a ellas. Este tema ha sido el centro de mi propio trabajo con los koalas, y volveré a hablar de él.

Un koala adulto pesa por término medio unos nueve kilogramos, y su cuerpo tiene unos 62 centímetros de longitud. Su pelaje es grueso y lanoso, sus patas largas y sus dedos presentan fuertes garras. En cada pie anterior los dos dedos más internos pueden oponerse a los demás como dos pulgares, al igual que el dedo más interno de cada pie posterior. Esta articulación hace del koala un buen trepador. Sobre el suelo, al que normalmente sólo desciende para desplazarse de un árbol a otro, el koala va a cuatro patas. La mayor parte del tiempo lo pasa subido a un árbol, durmiendo y buscando su comida entre las hojas, de manera intermitente, tanto durante el día como por la noche.

Los koalas suelen reproducirse cada dos años. La época de apareamiento se extiende desde principios de primavera hasta mediados de verano, y el período de gestación es de unos 35 días. Un koala recién nacido (generalmente la cría es única) tiene unos 19 milímetros de longitud y pesa unos 5,5 gramos. Esta minúscula criatura trepa sin ayuda hasta el marsupio, que es insólito porque se abre por la parte de atrás, y se agarra a un pezón. Unos seis meses más tarde, cuando ha alcanzado una longitud de unos 20 centímetros y le ha crecido un buen pelaje, el joven koala sale de la bolsa y a partir de entonces, y durante otros seis meses, es llevado sobre el dorso de su madre. A los seis meses de edad se desteta de las heces emulsionadas de la madre y está ya listo para su dieta adulta de hojas de eucaliptos.

Para los mamíferos, una dieta a base de hojas es una fuente de energía más pobre de lo que serían muchos otros tipos de alimentos. Además, los mamíferos son menos eficientes que otros animales en la digestión de hojas. Se considera por esta razón que muchos mamíferos que son folívoros arborícolas viven cerca del límite de su presupuesto energético. Por lo tanto, aun cuando el koala tiene éxito desde el punto de vista biológico, el fundamento de tal éxito es precario.

La cuestión del papel de los aceites volátiles de las hojas de eucalipto en el comportamiento alimentario del koala ha ocupado mucho tiempo y esfuerzo. La frecuente observación de que la preferencia del animal varía de una especie de eucalipto a otra, e incluso de uno a otro árbol de la misma especie, ha conducido a la suposición, ampliamente aceptada, de que un koala escoge activamente las hojas sobre la base de su contenido en aceites. Y tiene un motivo particular para hacerlo. Las razones aducidas para esta elección incluyen la acción antimicrobiana de los aceites en el voluminoso ciego del tracto intestinal del koala y la "necesidad" del efecto termogenerador de los aceites para mantener la temperatura corporal.

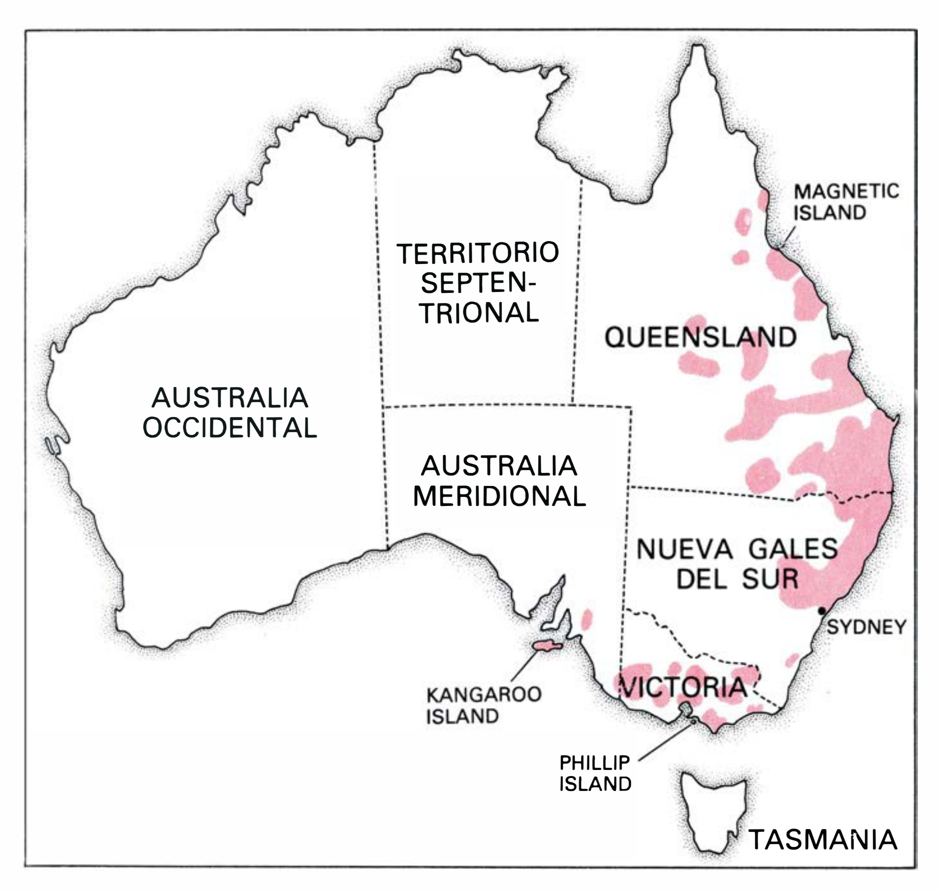
Ian Southwell, del Museo de Ciencias y Artes Aplicadas de Sydney, examinó recientemente la relación entre el koala y la composición de aceites volátiles en las hojas de eucalipto. No encontró correlación entre las variaciones en el nivel de los aceites volátiles de las hojas y las preferencias alimentarias de los koalas. Lo que su hallazgo demuestra es que el considerable éxito del koala en la ocupación del hábitat del eucalipto se debe a una capacidad general para detoxificar aceites que son perjudiciales para otros animales. En otras palabras, el koala es un marsupial que ha sido capaz de vencer el mecanismo de defensa química de los eucaliptos.

De esta manera el koala ha conseguido acceder a un recurso abundante. ¿Cómo lo utiliza el animal? ¿Hasta qué punto son adecuados el alimento, el agua y el cobijo que proporciona un eucalipto?

El primer paso para responder a estas preguntas es considerar la anatomía y la fisiología del sistema digestivo del koala. En común con otros mamíferos herbívoros, el koala es incapaz de digerir la celulosa, por lo que depende de microorganismos que la digieran. La localización de los microorganismos en el tubo digestivo proporciona una base para clasificar a los mamíferos herbívo-



KOALA RAMONEANDO en su hábitat característico, un eucalipto. Los koalas subsisten a base de las hojas de 35 de las aproximadamente 600 especies del género *Eucalyptus* que crecen en Australia. Las hojas proporcionan todos los nutrientes que el koala necesita, y por lo general la humedad suficiente, de manera que el animal no precisa beber. El koala se alimenta y duerme intermitentemente en los árboles, tanto de día como de noche.



DISTRIBUCION DEL KOALA, indicada por las zonas de color en este mapa. Antaño se cazaba tan ferozmente al animal por su piel que estuvo amenazado de extinción. Con la puesta en vigor de leyes de protección, Australia hizo también esfuerzos para restablecer la especie, en particular en Victoria, donde se instalaron colonias en las islas y en los bosques del continente. La colonia más septentrional parece haberse desarrollado a partir de un grupo de estos marsupiales introducidos en Magnetic Island.

ros en digestores pregástricos y postgástricos. En la categoría de los pregástricos se cuentan mamíferos euterios (placentados), el ganado por ejemplo, y marsupiales, como los canguros y los ualabíes; el grupo de los postgástricos incluye mamíferos euterios, tales como el caballo y los conejos, y marsupiales, como los falangeros zorro (*Trichosurus vulpecula*) y el koala.

Entre los digestores postgástricos, el lugar más común para los microorganismos es el ciego, una expansión del intestino posterior en la unión entre el intestino delgado y el grueso. En el ciego reside la característica más notable del sistema digestivo del koala. Observaciones recientes han establecido que ocupa cerca del 20 por ciento de la longitud del intestino postgástrico. De ese modo, el koala posee una cámara de fermentación en la que el paso de las hojas puede retardarse para que los microbios digieran la celulosa.

La posesión de un ciego tan grande significa que el requerimiento dietario del koala en carbohidratos probablemente se satisfaga en todo momento a través de la digestión microbiana de la celulosa. La posición del animal con respecto al nitrógeno, es decir, a las proteínas, está menos asegurada. Con el fin de determinar la capacidad del koala para permanecer en un equilibrio nitrogenado positivo, mis colegas y yo, en la Universidad de Nueva Gales del Sur, en Australia, mantuvimos enjaulado a un grupo de koalas, a los que alimentábamos exclusivamente con hojas frescas de *Eucalyptus punctata,* utilizado como alimento por los koalas. Medimos el nitrógeno incorporado y excretado en verano e invierno.

Como podía esperarse, encontramos que con un régimen de este tipo, el koala era capaz de permanecer en un balance nitrogenado positivo durante todo el año. Menos predecible era la forma de lograr ese equilibrio. Aunque la entrada de nitrógeno digerible parecía ser aproximadamente la misma en todo momento, la entrada a través de la dieta fue significativamente mayor en invierno. En otras palabras, los koalas tenían que consumir más hojas en invierno para permanecer en equilibrio de nitrógeno.

La explicación reside probablemente en diferencias estacionales en la calidad de las hojas. Los eucaliptos como *E. punctata* crecen rápidamente en primavera y a principios de verano, pero no demasiado una vez han florecido. Por ello es probable que las hojas de invierno sean más viejas y fibrosas y contengan menos nitrógeno digerible. He aquí una pequeña prueba de que los mamíferos que son folívoros arborícolas pueden estar viviendo cerca de los límites de su presupuesto energético.

La precaria naturaleza del presupuesto energético del koala se pone todavía más de relieve por el fenómeno conocido como "enfermedad de agotamiento". En épocas de sequía hay koalas que se vuelven comatosos y mueren con el estómago lleno. Sospecho que el factor principal es una deficiencia de nitrógeno, puesto que durante una sequía a los eucaliptos no les crecen hojas, o sólo unas pocas. Las hojas restantes son comparativamente viejas y contienen cada vez menos nitrógeno digerible. Es probable que la respuesta de un koala a un tal cambio en la calidad de las hojas disponibles sea comer más hojas. Si la calidad de las hojas se deteriora lo suficiente, la cantidad de ellas que se precisa e incluso la capacidad del sistema digestivo del koala pueden convertirse en factores limitantes. Bajo estas circunstancias puede resultar físicamente imposible que un koala satisfaga sus requerimientos nutritivos.

El koala tiene asimismo fama de no beber agua nunca, o de hacerlo raramente. Este es un tema común que se repite a través de las distintas leyendas de los aborígenes australianos acerca del koala. Algunos de los nombres aborígenes para el animal, como *koobor,* significan "no bebe agua". Curiosamente, *koala,* un nombre aborigen del distrito de Hawkesbury River, cerca de Sydney, no tiene tal connotación.

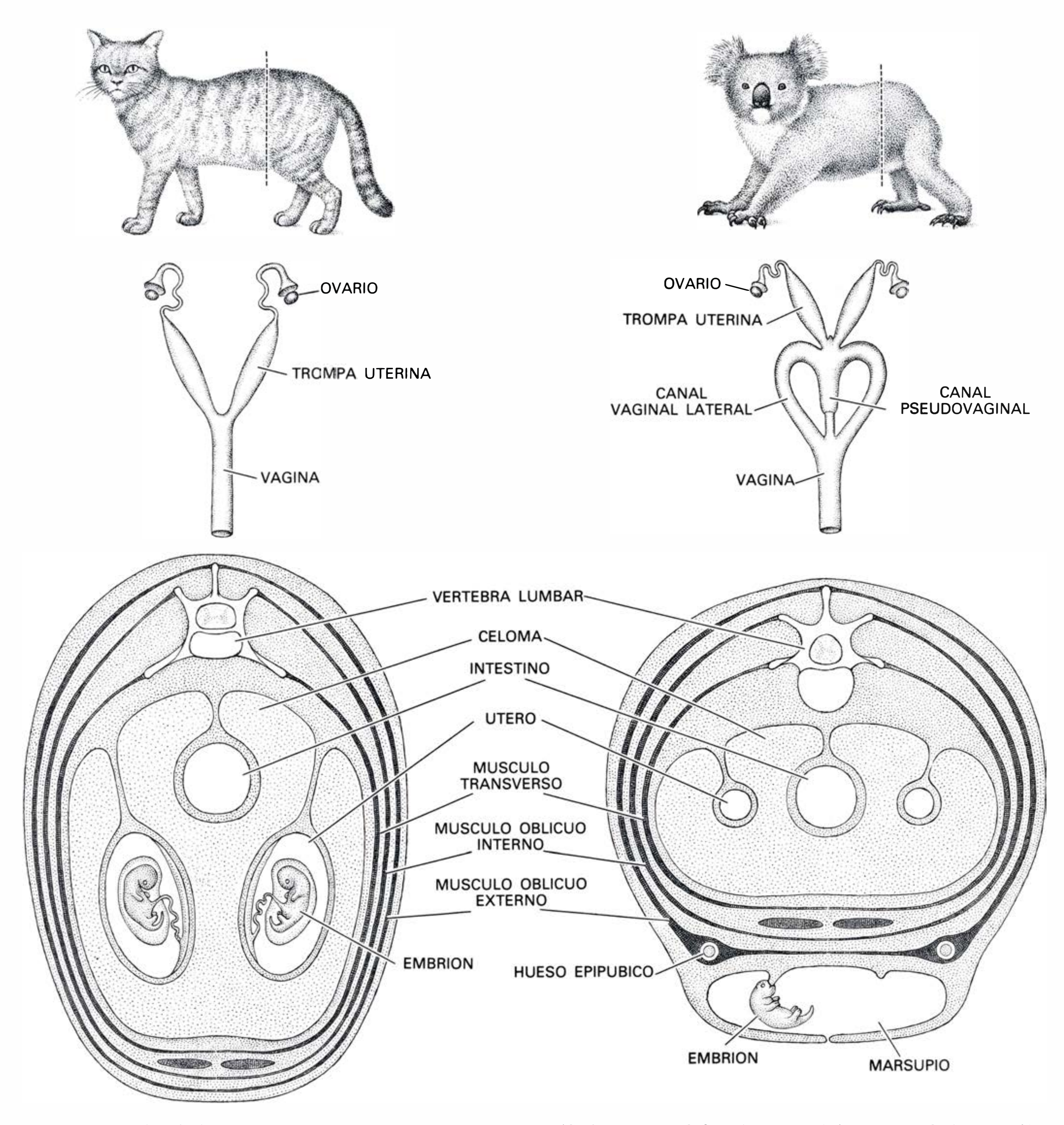
Las pruebas de que se dispone sugieren que, en condiciones normales, un koala obtiene el agua que necesita del rocío y de las hojas de eucalipto. En general, el agua se pierde en la orina, en las heces (ya sea como agua libre o como agua que se hubiera producido por la degradación metabólica del alimento) o por evaporación. La pérdida hídrica en la orina está regulada por el riñón. El contenido de las heces en agua libre está regulado por el intestino grueso, y el contenido de agua metabólica depende de la eficiencia digestiva. La pérdida por evaporación está íntimamente ligada a los mecanismos del animal para regular su temperatura corporal.

Comoquiera que se consiga el balance hídrico del koala, la tasa de renovación del agua es útil como una indicación de la disponibilidad y de la utilización del agua. Esta tasa puede determinarse (así como el contenido total de

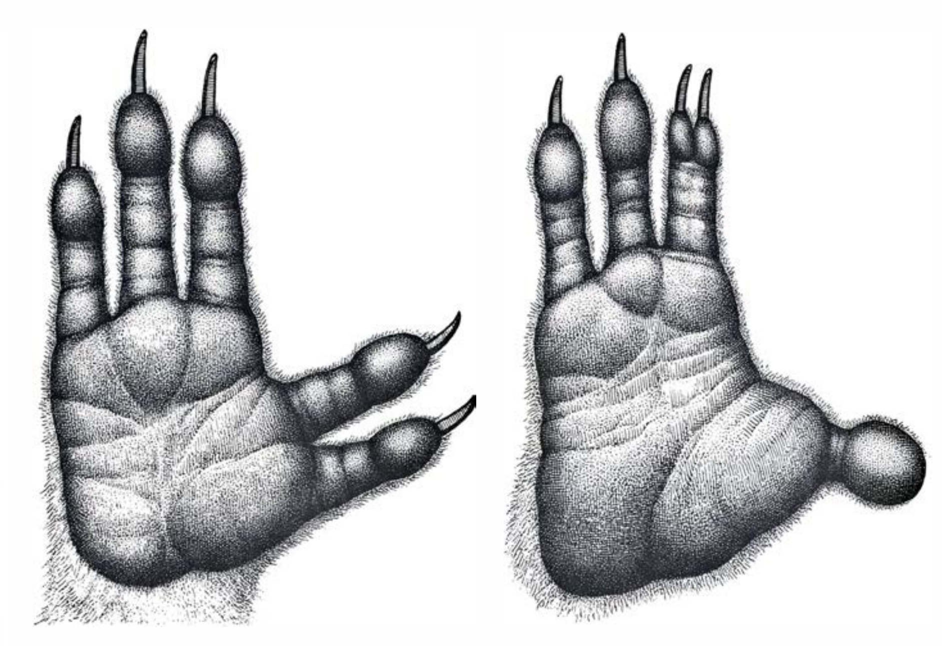
agua en el cuerpo) observando la dilución y la tasa de desaparición de agua marcada con tritio, el isótopo radiactivo del hidrógeno. Estudiamos el metabolismo del agua de koalas en libertad en tres poblaciones bien separadas: una en Magnetic Island, cerca de los límites septentrionales del área de distribución del koala, otra en Sydney, cerca del centro del área, y la tercera en Phillip Island, el límite meridional del área de distribución. No se encontraron diferencias significativas en el agua corporal total ni en la tasa de renovación.

El hallazgo indica que el microhábitat del koala es razonablemente uniforme en términos de requerimientos hídricos y de suministro de agua, con independencia de donde vivan los animales. El contenido en agua libre de las hojas de *Eucalyptus punctata* va desde el 40 por ciento (hojas viejas y fibrosas) hasta el 65 por ciento (hojas nuevas). Las hojas de otras especies de eucaliptos contienen por lo menos tanta agua como las de *E. punctata,* pues las hojas con un contenido hídrico inferior al 40 por ciento se secan y mueren. Puede concluirse que, en condiciones normales, las hojas de los eucaliptos proporcionan al koala alimento y agua en cantidades adecuadas.

El contenido de agua corporal del



SISTEMAS REPRODUCTORES comparados del gato (*izquierda*), un típico mamífero euterio (placentario), y del koala (*derecha*), un marsupial. Los dibujos de abajo muestran secciones transversales del cuerpo un poco por delante de las patas posteriores. El hueso epipúbico, distintivo de los marsupiales, ayuda a sostener la bolsa. Las crías de un mamífero placentario están mucho más desarrolladas al nacer que las de un marsupial. Un koala mide al nacer menos de 2 centímetros y prácticamente es todavía un embrión; se arrastra hasta el marsupio y se fija a un pezón. Permanece en la bolsa durante unos seis meses; para entonces, habrá crecido hasta una longitud de unos 20 centímetros y dispondrá de un abundante pelaje. Después de salir de la bolsa, la cría es transportada sobre el dorso de la madre durante otros seis meses. A los seis meses de edad puede alimentarse de hojas de eucalipto.



LAS MANOS Y LOS PIES del koala están adaptados al movimiento en los árboles: tienen garras largas y fuertes y dedos oponibles a los demás, como lo es el pulgar humano. La mano (*izquierda*) tiene dos de estos dedos, el pie uno. El segundo y el tercer dedo del pie, que están unidos, los emplea a menudo el animal para peinar su pelaje. Las palmas y las plantas de los pies tienen almohadillas granulares para agarrarse.

koala es relativamente alto (77,4 por ciento del peso del animal). Este valor se acerca al contenido hídrico del componente libre de grasas del cadáver de cualquier mamífero, una relación que sugiere que la ausencia total de reservas de grasa en los cadáveres de koala que he examinado es característica de este marsupial. La falta de grasa puede ser una consecuencia del precario equilibrio nutricio de este mamífero. También es lo que permite al animal poseer una proporción tan elevada de agua corporal. (Buena parte del agua se halla en el ciego, que puede contener gran cantidad de alimento húmedo.)

Cuando se examina de cerca la tasa de renovación de agua en el koala resulta clara la utilidad de transportar una considerable cantidad de líquido. Si esta tasa se expresa en función de la potencia cuatro quintos del peso del cuerpo, el efecto de variar el tamaño corporal se anula y se pueden hacer comparaciones inter- e intraespecíficas. Hemos comparado el koala y el bandicut hociquicorto *Isoodon macrourus,* un marsupial terrestre del tamaño de una rata que se alimenta de invertebrados. La tasa de renovación del agua de los bandicuts hociquicortos que viven en una isla sin agua potable es similar a la del koala (179 gramos por 0,8 kilogramos de peso corporal por día). En bandicuts que viven en el continente, rico en agua, la tasa es mucho mayor (243,8 gramos). Esta comparación refuerza la idea de que el alimento del koala incorpora un suministro de agua y que, en circunstancias normales, el animal no bebe. La idea se ve reforzada además por el hecho de que el koala tiene un riñón de tipo simple incapaz de un alto grado de conservación del agua. Los animales que han evolucionado para sobrevivir con una reducida ingesta de agua tienen típicamente un riñón que puede producir orina muy concentrada, con el fin de que en el cuerpo permanezca más agua disponible.

Investigamos el mecanismo por el que el koala logra el equilibrio hídrico efectuando mediciones en koalas enjaulados individualmente. A cada ejemplar se le proporcionaba diariamente una ración de ramas recién cortadas de *E. punctata.* En verano disponían siempre de agua para beber, pero en invierno se les retiraba a veces. Las entradas y salidas de agua se determinaron en forma de cambios en el peso.

La tasa de renovación del agua de los koalas enjaulados no fue superior a la mitad de la de sus congéneres en libertad. La razón más probable para esta diferencia es la reducción de actividad que impone la jaula. Como cabía esperar de un mamífero arborícola, las hojas eran la fuente principal de agua.

Aunque los koalas consumieron más alimento y agua en invierno, el agua que bebían contribuyó únicamente en cerca de una cuarta parte de la entrada hídrica, tanto en invierno como en verano. Resulta probable, pues, que el mayor ingreso de agua en invierno sea el resultado de una necesidad de consumir un volumen superior de hojas nutritivamente inferiores. En otras pa-

labras, los requerimientos nutricios dictarían el nivel de entrada de agua. No obstante, la incorporación de agua asociada con la comida permanece razonablemente constante, de 40 a 50 gramos por kilogramo de peso corporal por día. Por ello, y mientras las hojas que coma sean nutritivamente adecuadas, el koala puede obtener agua suficiente a partir de su dieta.

El koala pierde agua sobre todo a través de la evaporación desde sus superficies respiratorias. La pérdida de agua en la orina es el componente menos importante. La pérdida por evaporación y la producción de orina muestran una variación notablemente reducida de una a otra estación y cuando no dispone de agua para beber. La ligera disminución en la producción urinaria en ausencia de agua para beber sugiere que el koala posee un mecanismo de conservación de agua del tipo que se encuentra en el cerdo y el castor, cuyo riñón reabsorbe activamente urea cuando el agua para beber es limitada. Sin embargo, en el koala la reducción en el volumen de orina es tan ligera que sospecho que la urea se reabsorbe por razones nutritivas y no para la conservación del agua.

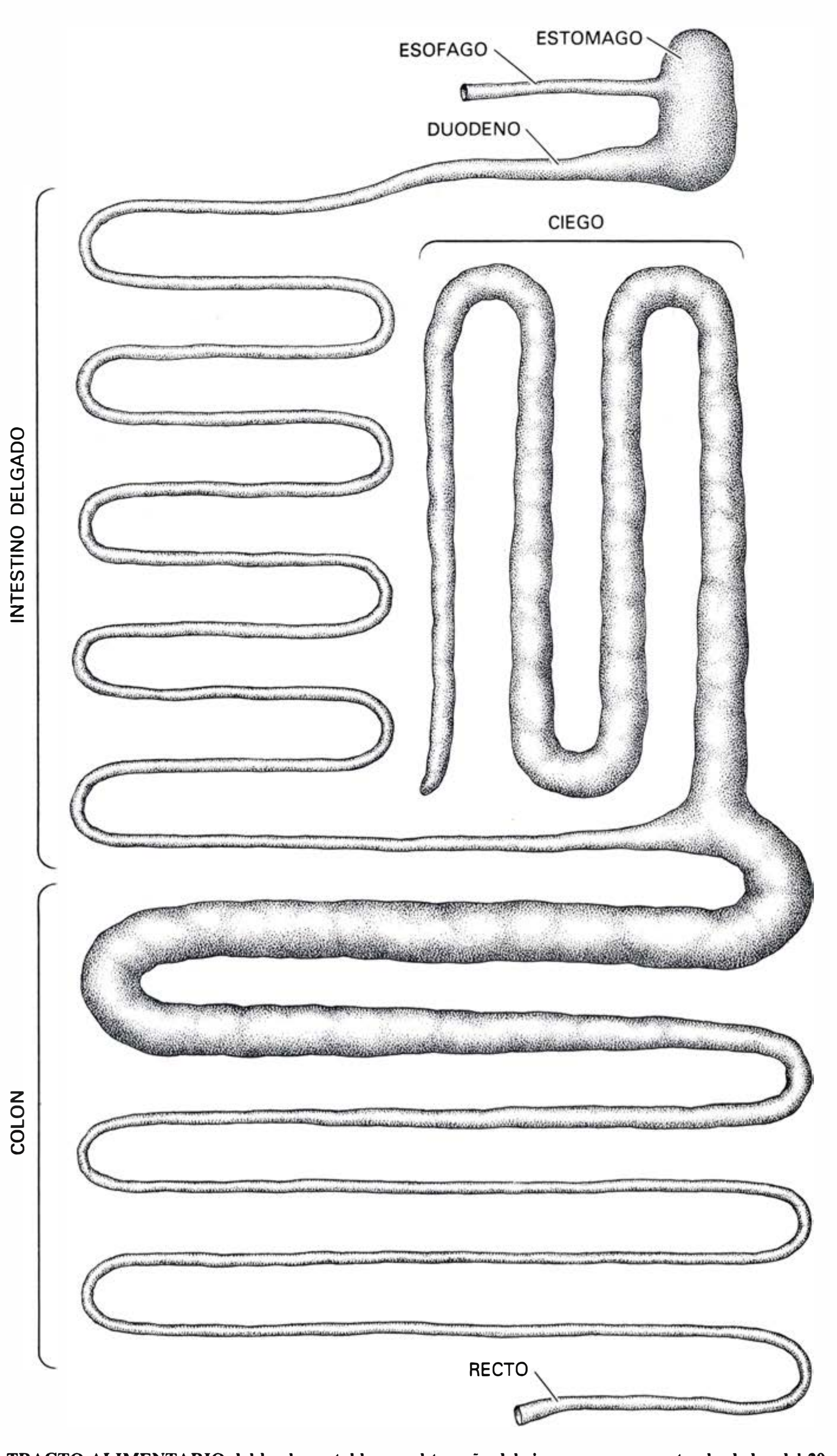
La pérdida de agua en las heces es probablemente el método más efectivo que tiene el koala para regular su balance hídrico. Las heces son siempre bastante secas. Cuando no se dispone de agua para beber, el contenido hídrico de las heces baja del 52 al 43 por ciento, que es el contenido hídrico de las heces de un camello deshidratado. Es evidente que el koala puede componérselas bastante bien con un suministro de agua disponible de forma regular, aunque no sea sobreabundante.

La elevada tasa a la que se pierde agua de las superficies respiratorias a través de la evaporación es una indicación de la importancia de la relación entre balance hídrico y balance térmico. Un examen de cómo regula el koala su temperatura corporal clarifica esta relación.

El koala se destaca de entre los marsupiales arborícolas porque no busca ningún tipo de refugio, una pauta de comportamiento (o, mejor, una falta de ella) que sólo se encuentra en otro grupo de marsupiales arbóreos, los canguros arborícolas. (Este comportamiento es bastante común entre los mamíferos placentarios que moran entre los árboles.) Por ello uno se siente incitado a descubrir qué tipo de protección frente a los extremos de las condiciones ambientales proporciona el pelaje del koala.

El pelaje dorsal, que es el más espeso, con 54,4 pelos por milímetro cuadrado, cubre el 77 por ciento de la superficie del cuerpo del animal. El pelo ventral tiene sólo la mitad de la densidad del pelaje dorsal y cubre el 13 por ciento de la superficie corporal. Las diferencias en densidad no tienen contrapartida en diferencias en la longitud del pelo, que es prácticamente la misma en todo el cuerpo tanto para los pelos protectores, más largos, como para la borra, más corta. Sin embargo, se encuentran diferencias estacionales en la longitud del pelo: la longitud de los pelos largos difiere más de la de los cortos en verano que en invierno.

El espeso pelaje dorsal es más oscuro



TRACTO ALIMENTARIO del koala, notable por el tamaño del ciego, que representa alrededor del 20 por ciento de la longitud del intestino. Puesto que el koala no puede digerir la celulosa de las hojas que constituyen su dieta básica y depende de microorganismos para hacerlo, el ciego sirve de cámara de fermentación en la que el paso de las hojas a través del tubo digestivo puede retardarse para que los microbios realicen su trabajo. El koala es capaz de detoxificar los aceites de las hojas de eucalipto.

que el ralo pelo ventral; por ello tiende a absorber el calor solar al tiempo que proporciona aislamiento. Los pelos ventrales esparcidos pueden erizarse, de manera que puede ajustarse la cantidad de aislamiento que proporciona el pelaje ventral. La combinación de estos tipos de abrigo dota al koala de un control ambiental que recuerda al del guanaco, animal sudamericano parecido a la llama, con una superficie dorsal densamente enmarañada (40 por ciento del área total) y zonas netamente definidas de piel casi desnuda en la superficie ventral. Cambiando su postura, el guanaco alcanza hasta una quintuple variación en su aislamiento térmico con el aire en calma, y cinco veces más que esto al viento.

Observaciones casuales de koalas subidos a árboles en días ventosos indican que, conforme la velocidad del viento aumenta, el koala tiende a arrollarse cada vez más en una bola compacta con el fin de presentar al viento únicamente su curvada superficie media dorsal. Los koalas hacen esto aun cuando la temperatura sea elevada. Si la velocidad del viento aumenta todavía más, doblan las orejas hacia delante, de manera que prácticamente nada se opone a la corriente de aire.

La idea de que cada koala lleva un refugio portátil se vio reforzada por mediciones que hicimos acerca de las características de aislamiento de las pieles de koala. (Nuestra muestra fue pequeña y, al ser accidental, se componía de más pieles de verano que de invierno; aunque no había mucha diferencia entre los dos tipos.) El pelaje dorsal del koala resultó tener el valor de aislamiento más alto entre los 12 marsupiales investigados hasta entonces. De hecho, se halla en el rango inferior de los valores que se obtienen para los animales árticos.

El efecto del viento sobre la capacidad de aislamiento del pelaje dorsal, denso y enmarañado, es reducido. Además, comparado con el pelaje de otros animales, el del koala muestra el descenso más pequeño en el valor de aislamiento al aumentar la velocidad del viento, por lo menos hasta unos 16 kilómetros por hora. El descenso medio fue del 14 por ciento, y el mínimo fue un notable 3 por ciento. Estos datos parecen indicar que el pelaje mantendrá un nivel importante de aislamiento a velocidades mucho más elevadas del viento y que, para un animal que vive en las copas de los árboles expuestas del bosque abierto, el pelaje proporciona una protección térmica excelente.

El efecto regulador de la temperatura que tiene el pelaje se complementa con la actividad metabólica. La tasa metabólica basal del koala es el 74 por ciento de la esperada para los marsupiales en general. (Entre los mamíferos placentarios, los folívoros como los perezosos y el poto muestran una divergencia similar de la tasa esperada.) El koala responde a temperaturas ambientales altas jadeando. A temperaturas ambientales bajas, la tasa metabólica reducida viene acompañada de un nivel elevado de aislamiento corporal total, contribuyendo el pelaje con el 50 por ciento (una proporción relativamente grande). El modelo de metabolismo-aislamiento del koala parece ser característico de muchos mamíferos arborícolas tropicales; quizá los marsupiales arborícolas en general difieran de los terrestres en este aspecto.

Al examinar la relación entre balance hídrico y balance térmico he empezado con el pelaje y he llegado a los mecanismos de termorregulación. Ahora puedo cerrar el círculo considerando la relación entre la pérdida de agua por evaporación (una característica de la termorregulación) y la producción metabólica de agua (un aspecto del equilibrio hídrico). Se calcula la cantidad de agua producida metabólicamente por cada gramo de oxígeno consumido sobre la base de la composición de nutrientes de las hojas y las heces. Se obtiene luego la relación entre producción metabólica de agua y pérdida de líquido por evaporación a partir de la relación entre consumo de oxígeno y cada uno de estos procesos. Estos cálculos han demostrado que las necesidades del koala para el enfriamiento por evaporación se satisfacen adecuadamente para una temperatura ambiente de hasta unos 30 grados Celsius, que raramente se sobrepasan durante mucho tiempo en el ambiente del animal.

En resumen, el koala es un animal cuyo nicho ecológico puede describirse (en el sentido más general) como el ambiente arbóreo que proporcionan los árboles que pertenecen principalmente al género *Eucalyptus*. Estos árboles suministran al animal alimento, agua y un lugar en el que vivir. Con sus sistema digestivo especializado, el koala puede vencer los efectos tóxicos de los aceites de los eucaliptos y puede extraer suficientes nutrientes y agua de las hojas de estos árboles. El modelo de termorregulación del koala está armonizado con este nivel de suministro de agua. Así (dejando aparte otras consideraciones), el animal es potencialmente capaz de medrar en la mayoría de bosques de Australia.

ADN superenrollado

En muchas formas del ADN, la doble hélice genera a su vez una hélice. Un sencillo modelo matemático describe y analiza este superenrollamiento, de importantes consecuencias en biología

William R. Bauer, F. H. C. Crick y James H. White

Imaginamos usualmente el ADN, el material genético de la mayoría de los organismos, como una doble hélice en la que dos cadenas de nucleótidos complementarios (cuya secuencia constituye el mensaje genético) se enrollan alrededor de un eje rectilíneo común. Pero a menudo el eje de la doble hélice no es recto, sino curvo. Incluso se puede enrollar en el espacio formando una nueva hélice de orden superior, y entonces hablamos de ADN superenrollado. Muchos ADN conocidos exhiben algún tipo de superenrollamiento, al menos durante una fase de su ciclo vital. La valoración de esta estructura y de sus consecuencias es, por tanto, esencial para comprender plenamente la biología del ADN.

El superenrollamiento se da en ADN de diferentes tipos y tamaños, incluso en algunos que no tienen estructura de doble hélice, y adopta varias formas distintas. Por ejemplo, en la cromatina (un agregado de ADN y proteínas) de los organismos superiores, el ADN está enrollado sobre un gránulo de proteína formando una superhélice solenoide sinistrorsa. Nos ocuparemos aquí sobre todo de otro tipo de superenrollamiento, que no necesita apoyarse sobre proteínas. Se trata del superenrollamiento del ADN circular cerrado: moléculas bicatenarias en las que cada cadena de nucleótidos forma una curva cerrada continua, sin interrupciones.

La existencia de estos anillos retorcidos de ADN se propuso en principio para explicar un descubrimiento sorprendente sobre el ADN de polioma, un pequeño virus que causa cáncer a los ratones. En 1963 se publicó que, al centrifugar una suspensión de este ADN, se separaban tres componentes, distinguibles por su velocidad de sedimentación. Los tres componentes, llamados I, II y III en orden de velocidad de sedimentación decreciente, tenían el mismo peso molecular, según se desprendió de ensayos ulteriores. Debía haber, por tanto, alguna diferencia estructural que explicara las diferentes velocidades de sedimentación. Es decir, las moléculas debían de tener formas distintas.

La microscopía electrónica confirmó esta suposición. Las moléculas del componente III eran sin duda lineales, o sea, abiertas, mientras que las de los otros dos componentes eran circulares, o sea, sin extremos visibles. El descubrimiento de que los componentes I y II eran circulares no zanjó la cuestión, pues no se sabía en qué se diferenciaban. En 1965, Jerome Vinograd, del Instituto de Tecnología de California, resolvió ingeniosamente el misterio, introduciendo el concepto de superenrollamiento del ADN.

Vinograd sugirió que las dos cadenas de un ADN circular cerrado (una molécula circular cuyas dos cadenas de polinucleótidos están completamente intactas) dan usualmente menos vueltas una alrededor de la otra que las de una molécula de ADN lineal de la misma longitud. En otras palabras, en la molécula circular cerrada del componente I hay menos giros helicoidales que en las moléculas de los componentes II y III. El componente I da la impresión de haber sufrido, antes de unirse las terminaciones de la doble hélice, un cierto número de giros de 360 grados en sentido inverso al del enrollamiento normal de la hélice.

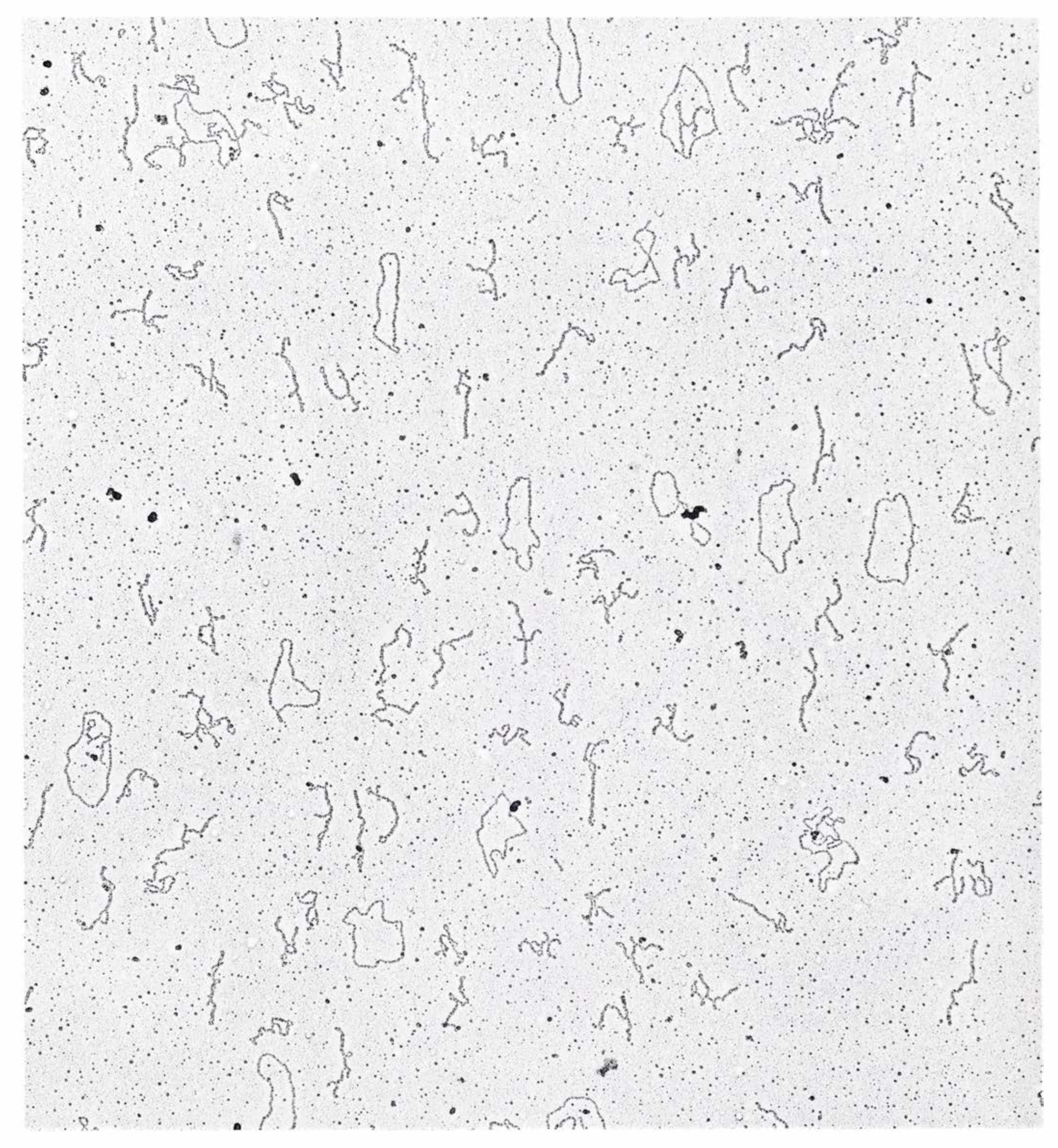
La diferencia entre los componentes I y II del ADN de polioma es que las moléculas circulares de este último no son cerradas: cada molécula tiene, por lo menos, una incisión o rotura en una de sus cadenas. Por el contrario, las dos cadenas del ADN circular del componente I son curvas cerradas y dan menos vueltas helicoidales de lo normal. Las fuerzas que estabilizan la doble hélice son lo bastante fuertes para resistir este estiramiento y, como veremos más adelante, lo compensan formando superhélices. Las moléculas circulares cerradas del componente I resultan más compactas que las discontinuas del componente II y por eso sedimentan más aprisa.

El superenrollamiento es un fenómeno frecuente, típico de muchos de los ADN de mayor interés médico y biológico, como los de un grupo entero de virus oncogénicos, que incluye los poliomas y los virus de los papilomas (verrugas) humanos. El ADN de las mitocondrias humanas y de otros animales también está superenrollado. Los animales poseen, además, gran número de pequeñísimas moléculas superenrolladas de ADN, de función desconocida. Llama particularmente la atención que la mayoría de los menores genomas (dotaciones génicas) conocidos caigan en esta categoría, entre ellos los responsables de la fertilidad y la resistencia a antibióticos en las bacterias. Además, la integración física de un pequeño fragmento de ADN en una molécula mayor parece requerir que el elemento que se integra esté superenrollado. Así son los vectores, o vehículos, usados para el trasplante de genes en ingeniería genética. Quizá lo más notable sea que los virus oncogénicos de ARN, como el de Epstein-Barr, emplean a menudo la transcriptasa inversa para hacer copias de ADN que se superenrollan en el curso de la infección. Es interesante señalar que se ha observado superenrollamiento en moléculas que contienen desde 350 hasta 1.750.000 nucleótidos en cada cadena.

Prácticamente todas las propiedades físicas, químicas y biológicas del ADN, como su transcripción, comportamiento hidrodinámico, energética y enzimología, resultan afectadas por la circularidad y el superenrollamiento. La comprensión del mecanismo de superenrollamiento y sus consecuencias exige resolver problemas de considerable complejidad matemática. Afortunadamente, dos ramas de la matemática facilitan sustancialmente este esfuerzo: la topología, que estudia las propieda-

des de las estructuras que se conservan al deformarlas, y la geometría diferencial, que aplica los métodos del cálculo diferencial al estudio de curvas y superficies. Describiremos a continuación un modelo matemático del ADN circular cerrado y discutiremos sus implicaciones para el ADN real. Trataremos de algunos métodos ideados para medir el superenrollamiento y sus efectos.

Consideremos ante todo la estructura física de la doble hélice: consta de dos cadenas de azúcar y fosfato a las que se unen las bases nucleotídicas. Las bases de las dos cadenas opuestas están apareadas y cada giro completo de la hélice abarca unos 10 pares de bases (el número exacto depende de la configuración precisa de la molécula). Para estudiar matemáticamente el superenrollamiento es aconsejable construir un modelo que represente el ADN como una cinta estrecha y retorcida, de grosor infinitesimal. La manera más obvia de construir el modelo es hacer coincidir los bordes de la cinta con las cadenas de azúcar y fosfato del ADN. Al llevar a la práctica este modelo para el caso de una molécula bicatenaria lineal, se advierte lo erróneo de la



MOLECULAS SUPERENROLLADAS DE ADN observadas al microscopio electrónico. Cada una es un anillo de ADN bicatenario. En los anillos extendidos, una de las cadenas de la doble hélice está abierta, o rota, de modo que el anillo queda relajado, no superenrollado. Los anillos retorcidos no tienen roturas. Su ADN presenta usualmente menos vueltas de hélice que un ADN lineal de la misma longitud. La tensión introducida por el defecto de vueltas se compensa, si no hay cortes, por superenrollamiento. Estas moléculas son plásmidos extraídos de la bacteria *Escherichia coli*, separados de otras moléculas de ADN de *E. coli* por centrifugación. Aparecen aumentadas unas 25.000 veces. La micrografía fue tomada por Gary Cohen.

elección, pues la línea media de la cinta, equidistante de ambos bordes, da vueltas alrededor del eje rectilíneo de la doble hélice. En nuestro estudio es preferible un modelo cuya línea media coincida con el eje de la doble hélice. Exigimos, además, que la cinta sea siempre perpendicular a las pseudodiadas, o ejes binarios de rotación, distribuidas a lo largo de la doble hélice [*véase la ilustración inferior*]. (Cada par de nucleótidos define un eje y el espacio entre pares sucesivos define otro; todos estos ejes son perpendiculares al eje de la doble hélice y su situación es independiente de la secuencia de bases de la molécula.)

La cinta sigue el eje de la doble hélice y se retuerce con las vueltas de las dos cadenas. Asignamos además orientaciones contrarias a los bordes de la cinta, porque las secuencias de átomos de las dos cadenas tienen sentidos opuestos; (no importa el sentido que asignemos a un borde, con tal que sea opuesto al otro). Se pueden estudiar matemáticamente varios aspectos de este modelo, pero nos ocuparemos principalmente de la relación entre los dos bordes de la cinta.

Al unir los extremos de la cinta, cada borde describe una curva cerrada en el espacio tridimensional. Además, si la cinta representa una molécula circular cerrada de ADN, se retuerce cierto número de veces y las dos curvas descritas por sus bordes quedan enlazadas, de modo que no se pueden separar sin "cortar" una de ellas. Si cada curva representa una molécula formada por enlaces covalentes, como en el caso de la doble hélice, se dice que las dos moléculas están unidas por un enlace topológico. Este enlace tiene la peculiaridad de que no hay enlaces covalentes entre las dos moléculas, pero habría que romper enlaces covalentes para separarlas. (El concepto de enlace topológico fue introducido en química por Edel Wasserman y sus colegas, de los laboratorios Bell.)

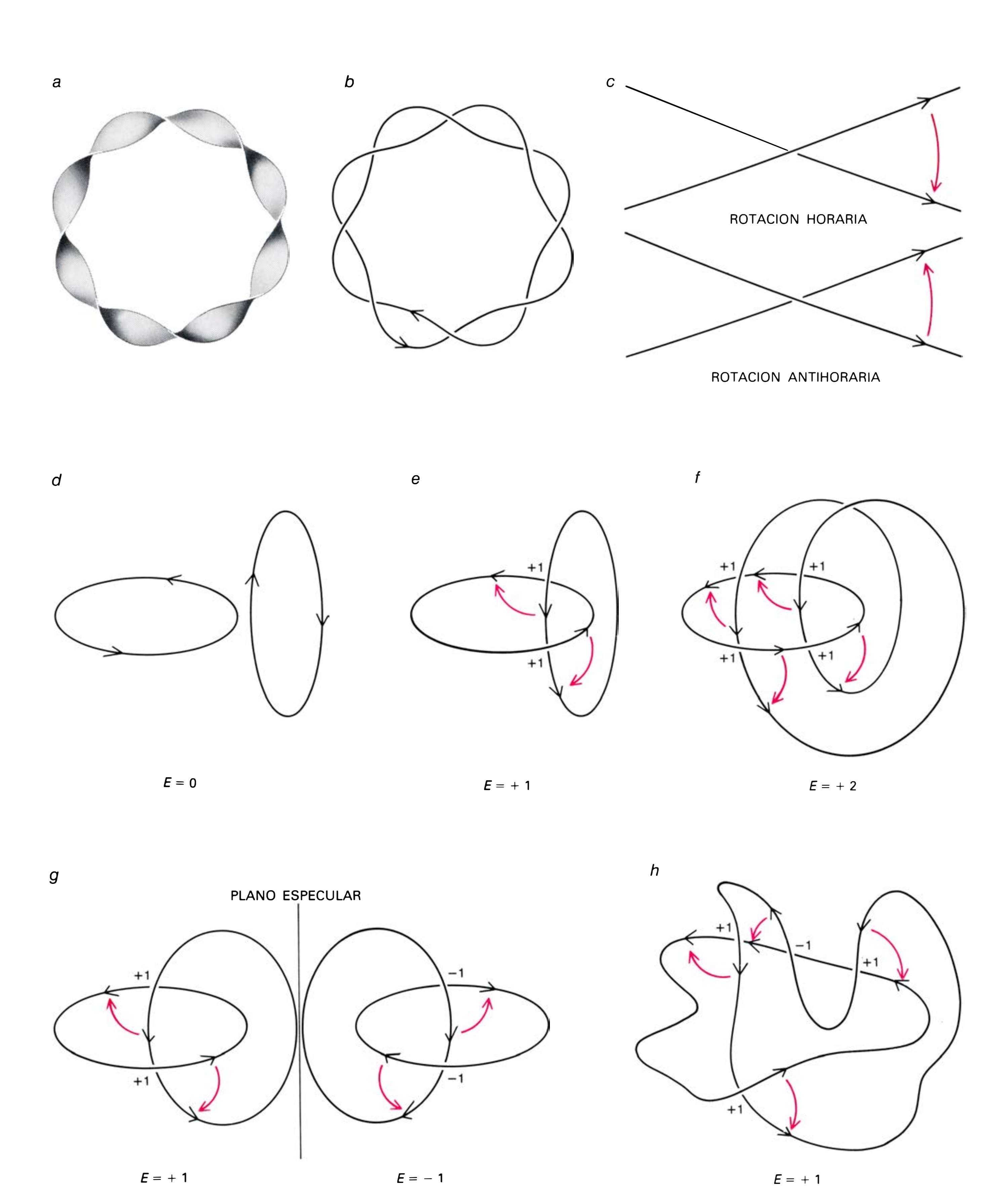
En términos matemáticos, el enlace de dos curvas cerradas es una propiedad topológica: el enlace permanecerá inalterado por mucho que se deformen, estiren o retuerzan las cadenas, en tanto no se rompa alguna de ellas. Es útil concebir un valor numérico que describa el enlace entre las dos curvas que representan a las dos cadenas del ADN. Parece razonable definir este índice de enlace E de modo que valga 0 en el caso de dos curvas cerradas no enlazadas, 1 si una curva se enrolla alrededor de la otra una vez, 2 si lo hace dos veces, y así sucesivamente. Como hemos atribuido orientaciones a los dos bordes de la cinta, es preferible definir el índice de enlace de modo que tenga signo positivo o negativo según la orientación de las curvas. Hay varias formas de calcular el índice de enlace; una de las más convenientes consiste en fijarse en los cruces de las dos curvas en una proyección o representación bidimensional del modelo [*véase la ilustración de la página opuesta*]. Puede asignarse un valor de $+1$ o -1 a cada cruce según la dirección en que haya de girar la curva superior para que coincida con la inferior. El índice de enlace se obtendrá sumando los valores de todos los puntos de cruce y dividiendo por 2 (número de curvas).

El índice de enlace E se define así como un entero positivo o negativo que describe una propiedad de dos curvas cerradas en el espacio. Para poder separar las curvas sin cortarlas, E tiene que ser 0, pero la recíproca no siempre es cierta. Si las curvas en cuestión son los bordes de una cinta cerrada retorcida N veces, su índice de enlace será $+N$ o $-N$, según el sentido de la rotación. Nótese que, aunque en el modelo de ADN los bordes de la cinta no coinciden con las cadenas de azúcar y fosfato de la doble hélice, el índice de enlace de la cinta es el mismo que el de las cadenas.

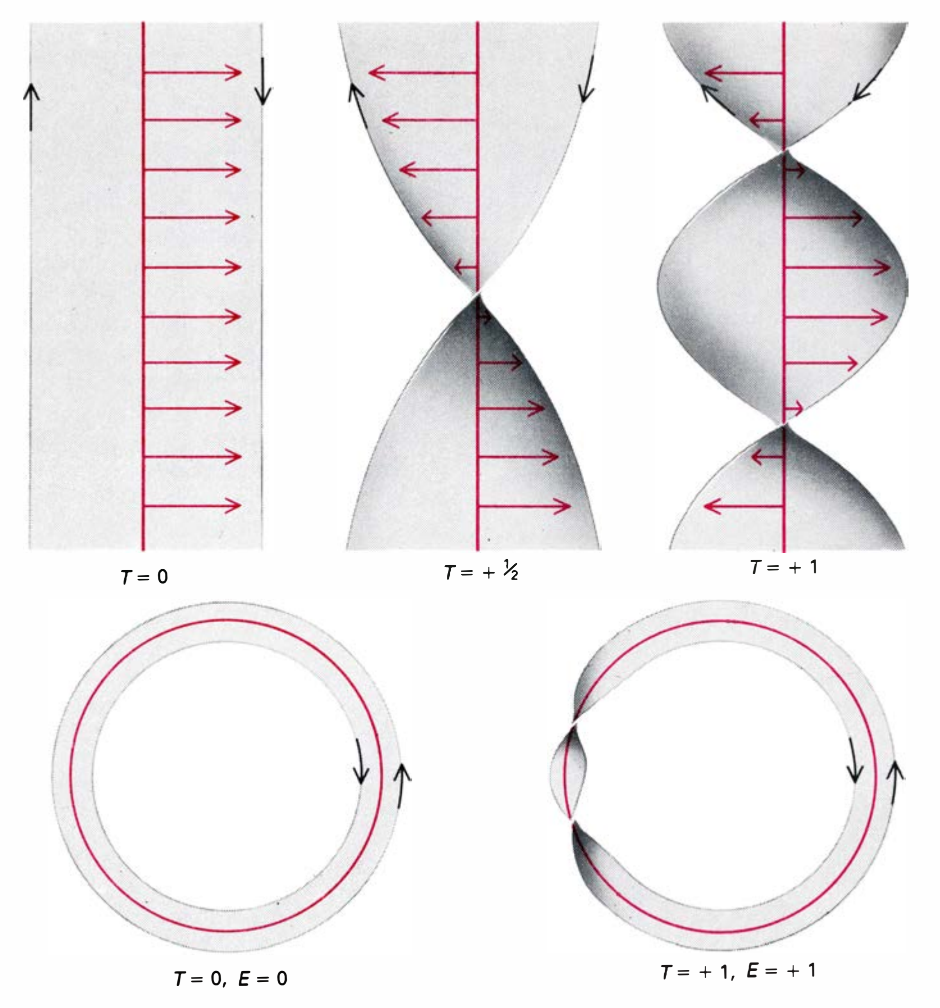
En una molécula de ADN circular, cerrada y relajada de 5000 pares de bases, con una vuelta de la hélice cada 10 pares de bases, el índice de enlace será $+500$. Convenimos en que este número sea positivo porque la doble hélice del ADN es dextrorsa. Se llama relajada a una molécula de ADN circular cerrada si su eje puede disponerse completamente en un plano. Como veremos, las moléculas superenrolladas naturales se pueden relajar de muchas maneras distintas. Una molécula circular discontinua está relajada en su estado natural, pero no se puede hablar estrictamente de índice de enlace, porque al menos una de sus cadenas está abierta.

MODELO DEL ADN SUPERENROLLADO basado en dos elementos estructurales de la doble hélice (*izquierda*): el eje (*recta gruesa de color*), a cuyo alrededor se enrollan las cadenas de azúcar y fosfato, y la serie de ejes binarios (*recta fina*), definidos por las bases unidas a las cadenas. Las bases, cuya secuencia constituye el mensaje genético, están apareadas; cada par de bases define un eje y otro el espacio entre bases sucesivas (el dibujo muestra sólo un par de bases). El eje binario definido por cada par de bases es la línea perpendicular al eje de la hélice alrededor de la que habría que rotar una de las bases para llevarla a la posición de la opuesta. El modelo elaborado de acuerdo con estas características (*derecha*) es una cinta estrecha e infinitamente delgada cuyo eje coincide con el eje de la doble hélice y cuya superficie es siempre perpendicular a los ejes binarios definidos por los pares de bases de la doble hélice. La cinta sigue las curvaturas del eje del ADN y se retuerce siguiendo las vueltas descritas por las dos cadenas.

Otro modo de analizar este modelo del ADN es observar la torsión de la cinta, en vez de la relación entre sus bordes. Si el eje de la cinta es una línea recta, es fácil intuir un valor numérico que exprese la torsión. Asignaremos un valor de $+1$ por cada vuelta de 360 grados a la derecha y un valor de -1 si



ADN CIRCULAR CERRADO representado por una cinta retorcida con sus extremos unidos (*a*). Como las secuencias de átomos de las dos cadenas de la doble hélice tienen sentidos opuestos, conviene asignar a los bordes de las cintas también sentidos contrarios. Los bordes de la cinta son curvas cerradas enlazadas matemáticamente una con otra en el espacio tridimensional (*b*); en otras palabras, no se pueden separar sin romper al menos una. Esta relación se puede describir matemáticamente por un índice de enlace E que refleja el número de veces que una curva pasa por el anillo de la otra y cuyo signo depende de los sentidos atribuidos a las curvas. Un modo de calcular E es examinar una proyección de las curvas y asignar un índice a cada punto de cruce de una curva sobre otra (pero no de una curva sobre sí misma) de acuerdo con la siguiente regla: + 1, si se requiere una rotación en sentido horario para hacer coincidir la curva superior con la inferior, y − 1, para la rotación opuesta (*c*). (Esta regla es la inversa de la usual en matemática, pero asegura que el índice de enlace del ADN circular cerrado y dextrorso sea positivo.) Para obtener E se suman los índices y se divide por 2 (número de curvas). El índice de enlace así obtenido es un número entero, igual a 0 si las dos curvas no están entrelazadas (*d*), + 1 o − 1 si una curva pasa por el anillo de la otra una sola vez (*e*), + 2 o − 2 si dos veces (*f*), etcétera. El signo de E cambia si varía la orientación de cualquiera de las curvas o si se reflejan las curvas en un espejo (*g*). El índice de enlace no cambia por mucho que se deformen las curvas (*h*). Puesto que el índice de enlace de una cinta retorcida es igual al de las cadenas de la molécula de ADN que representa, expresa una invariante importante de la posible estructura superenrollada del ADN.



TORSION DE UNA CINTA. Puede calcularse colocando una flechita en ella, perpendicular a su eje y orientada hacia uno de los bordes. Dado que al desplazarse la flecha a lo largo de la cinta gira alrededor de su eje, podemos definir la torsión total *T* como la integral de la velocidad angular de rotación respecto a la longitud del arco de curva descrito sobre el eje. El signo positivo o negativo de esta cantidad depende de que la rotación de la flecha alrededor del eje sea dextrorsa o sinistrorsa. Cuando el eje de la cinta está contenido en un plano, como aquí, la torsión total es fácil de calcular, porque equivale al número de vueltas que da la flecha alrededor del eje. Se puede calcular separadamente la torsión en diferentes partes de la cinta y sumar para obtener el total. La torsión, a diferencia del enlace, no siempre es un entero.

es a la izquierda. La definición de torsión no resulta tan obvia cuando el eje de la cinta no es recto. Quizá el mejor modo de comprender el concepto de torsión sea imaginar una flechita perpendicular al eje de la cinta, que apunte hacia uno de los bordes [*véase la ilustración de esta misma página*]. Al desplazar la flecha a lo largo de la cinta, gira alrededor del eje. La torsión puede definirse como la integral de la relación entre la rotación angular de la flecha y la longitud del arco recorrido sobre la curva eje.

En el caso especial de que el eje de la cinta esté contenido en un plano, la torsión se puede medir sencillamente por el número de vueltas que da la flecha alrededor del eje al desplazarse a lo largo de la cinta. Por ejemplo, si la cinta representa una molécula de ADN de 5000 pares de bases, circular, cerrada y relajada (es decir, con su eje en un plano), la flecha completa una rotación cada vuelta de la doble hélice y por tanto la torsión total T es + 500, con el signo + porque la doble hélice es dextrorsa.

En este caso, por tanto, el índice de enlace y la torsión son + 500. Este ejemplo podría hacernos suponer que el índice de enlace y la torsión son siempre iguales, pero no es así. E importa mucho distinguir entre estas dos cantidades. Para comenzar, el enlace es una propiedad topológica y la torsión es una propiedad geométrica: al deformar la cinta se puede alterar su torsión. Además, para calcular el índice de enlace, que es siempre un entero, hay que tener en cuenta toda la cinta. Por el contrario, la torsión, que no tiene que ser un entero, se puede definir localmente y se pueden sumar las torsiones de diferentes segmentos para obtener el total.

Conviene advertir que este modelo no representa, ni siquiera aproximadamente, las propiedades mecánicas del ADN, entre otras cosas porque se atribuye un grosor infinitesimal a la cinta matemática. Una cinta real, con un grosor finito, sometida a las torsiones que nos ocupan, experimentaría muchas deformaciones locales que no estamos teniendo en cuenta. El modelo de cinta se ideó para resaltar el índice de enlace del ADN, valor cuántico insensible a esas pequeñas deformaciones.

La distinción entre enlace y torsión sugiere otra pregunta: ¿tiene significado geométrico la diferencia entre el índice de enlace y la torsión total de una cinta? Dio la casualidad de que cuando los bioquímicos empezaban a estudiar el ADN circular, los matemáticos se interesaron por los enlaces y torsiones de las cintas. En 1968 uno de nosotros (White) demostró que la diferencia entre el índice de enlace y la torsión total de una cinta dependía exclusivamente de la curva del eje de la cinta. (Esta cantidad es bien conocida por los matemáticos como la integral gaussiana de la curva eje.) En otras palabras, si los ejes de dos cintas cerradas siguen la misma curva en el espacio tridimensional, la diferencia entre sus índices de enlace y sus torsiones totales será exactamente la misma, aunque las cintas se retuerzan de modo totalmente distinto.

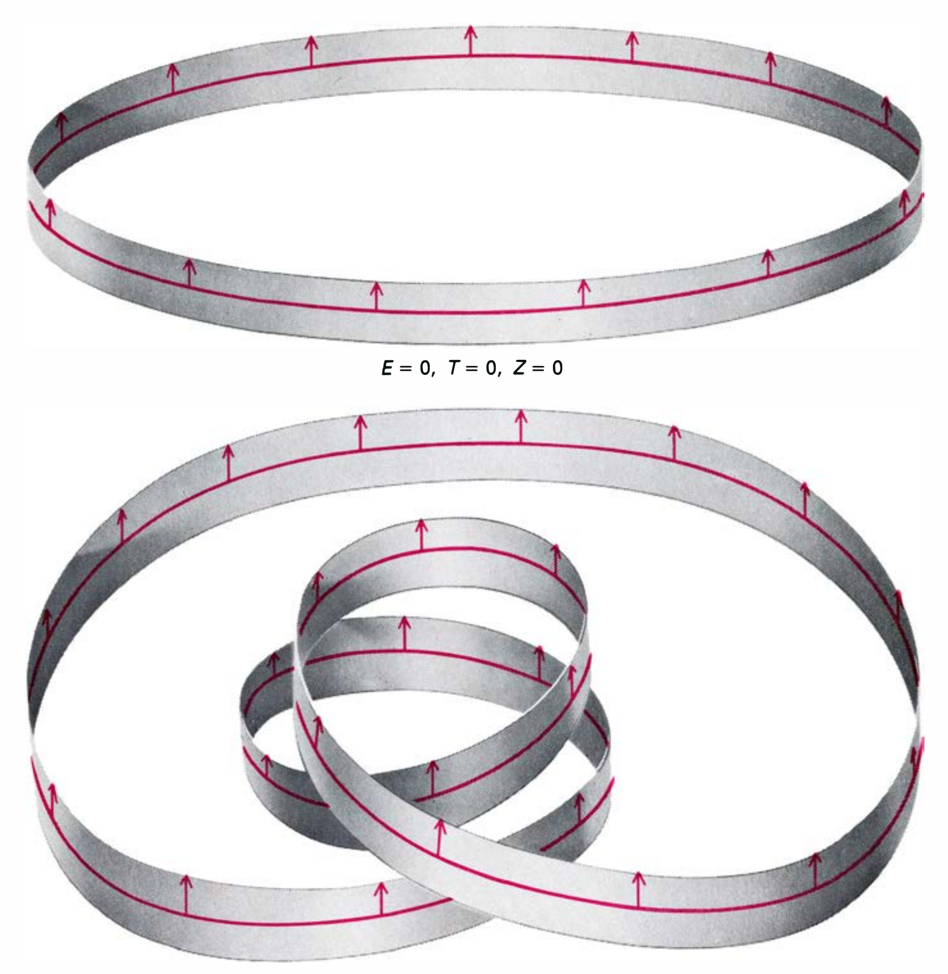
Por la misma época, Vinograd, que desconocía este resultado, pidió a F. Brock Fuller, un matemático del Instituto de Tecnología de California, que estudiara el superenrollamiento. Fuller, al ocuparse de la relación entre enlace y torsión, dio a su diferencia un nombre pintoresco, que traduciremos por enzarcillamiento. Es decir, el enzarcillamiento Z de una cinta cerrada en el espacio tridimensional es igual a la diferencia entre el índice de enlace E y la torsión total T. El enzarcillamiento de una cinta es un número de gran aplicabilidad, cuyo valor cambia, en general, si el eje de la cinta se deforma en el espacio. El enzarcillamiento, como la torsión, no es una propiedad topológica, sino geométrica.

El enzarcillamiento se puede obtener calculando la integral gaussiana, pero generalmente es mucho más sencillo calcular el índice de enlace y la torsión total y hallar su diferencia. Sólo en algunos casos especiales conviene calcular Z directamente. Por ejemplo, si el eje de una cinta está totalmente contenido en un plano o en la superficie de una esfera, se puede demostrar que Z es 0. Al sustituir este valor en la ecuación $Z = E - T$, resulta $E = T$, lo que explica que, en el ejemplo del ADN circular, cerrado y relajado, tanto

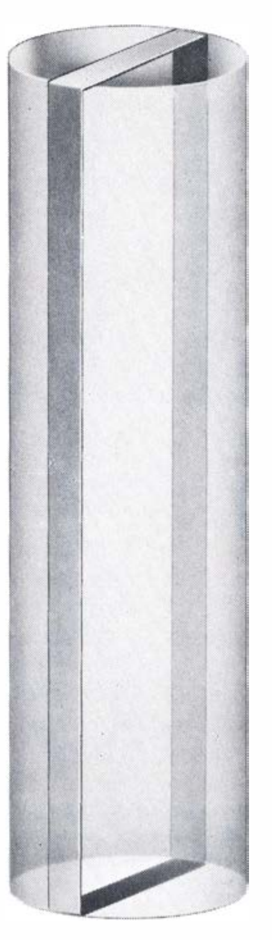
E como T fueran igual a + 500. Consideremos ahora lo que ocurre si el eje de este ADN se distorsionara de manera que su enzarcillamiento dejara de ser nulo. (No se puede estimar intuitivamente el enzarcillamiento de una curva; una cinta enroscada, en el sentido usual de la palabra, puede tener un valor de enzarcillamiento nulo.) Al variar el enzarcillamiento del ADN, el índice de enlace permanece constante, ya que sólo puede alterarse si se rompe alguna de las cadenas de la doble hélice, y, por tanto, debe variar la torsión. Esta relación es la base del superenrollamiento.

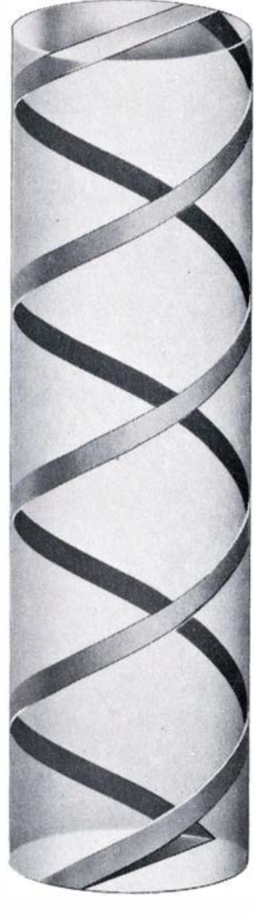
El índice de enlace, la torsión total y el enzarcillamiento de una cinta no dependen de la situación u orientación de la cinta en el espacio. También son independientes de la escala; pero si se invierte uno de los ejes del espacio (como si se reflejara la cinta en un espejo) o se invierten los tres ejes (tomando la figura simétrica respecto a un punto), las tres cantidades cambian de signo. Por el contrario, si se invierten dos ejes cualesquiera, como si miráramos a través de un microscopio ordinario (no un microscopio de disección), los signos no cambian. Hay otros dos casos de interés práctico. Al hacer una copia por contacto de una proyección de la cinta, cambian los signos de las tres cantidades. En cambio, si el negativo de una fotografía de esta proyección se coloca correctamente en una ampliadora, los signos de las tres cantidades de la imagen resultante no habrán cambiado. Por el contrario, si se voltea el negativo al ponerlo en la ampliadora, cambiarán los signos de las tres cantidades de la imagen resultante.

Resumiendo, cualquier operación que cambie el sentido de dextrorso a sinistrorso, sin introducir otras distor-

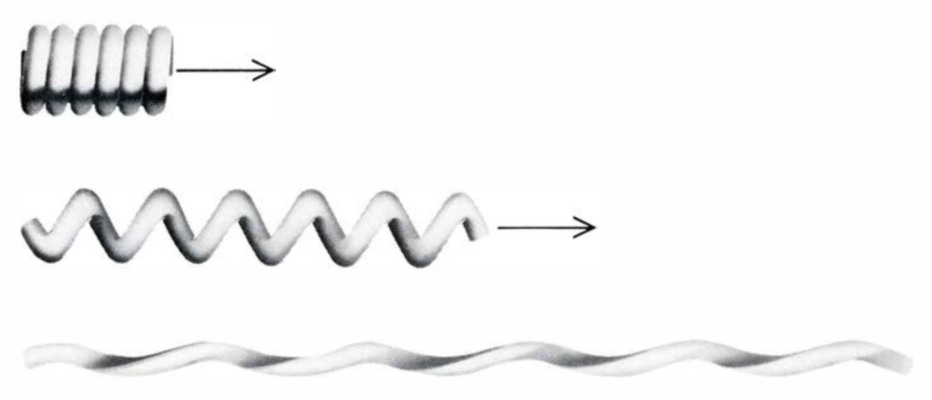




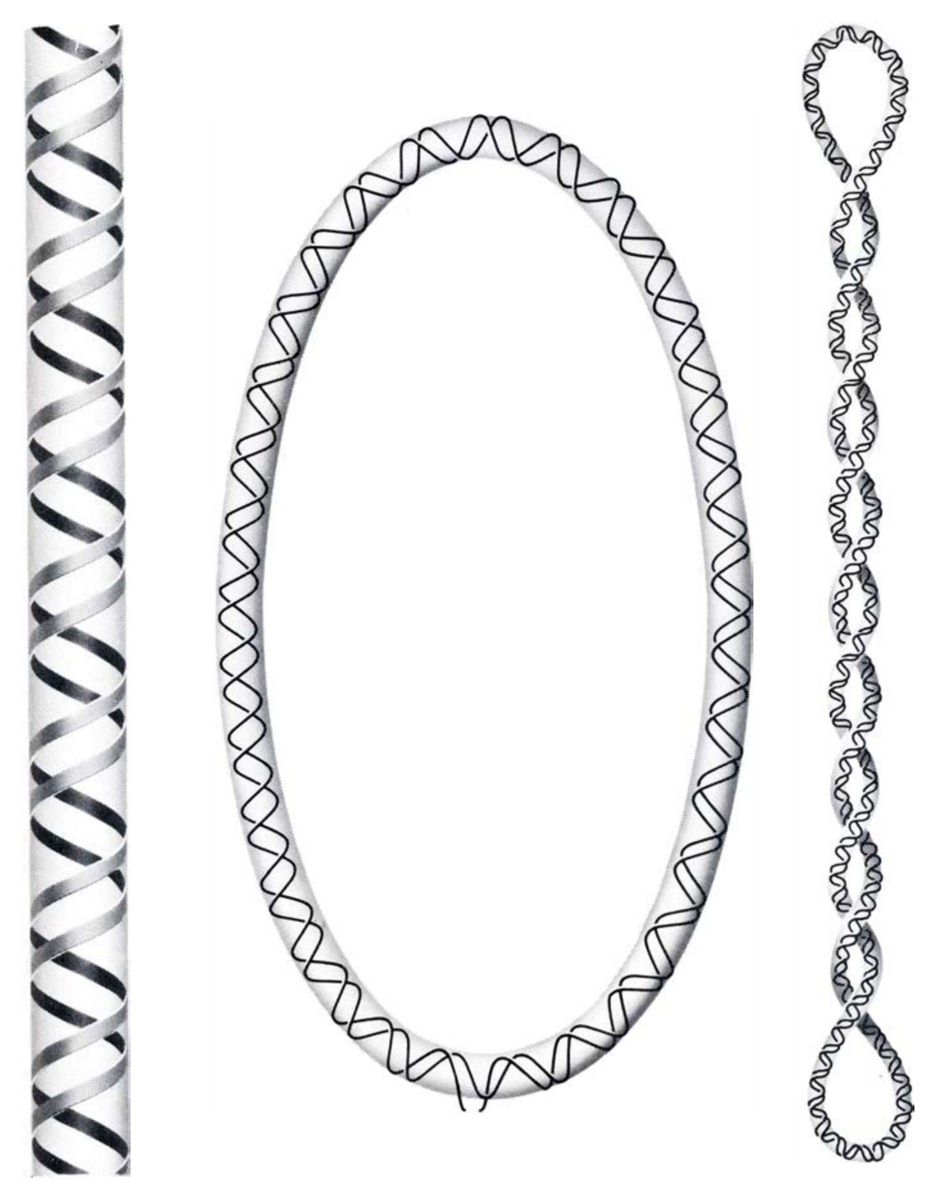




INDICE DE ENLACE y torsión total de una cinta. Uno y otra son iguales cuando el eje de la cinta está contenido en un plano, como aparece arriba. Las dos cantidades no son siempre iguales, según muestra la configuración del centro, cuyo índice de enlace es un entero positivo, pero la torsión total es próxima a 0. La diferencia entre el índice de enlace y la torsión total de una cinta se llama enzarcillamiento, Z, de modo que $Z = E - T$. El enzarcillamiento depende exclusivamente de la curva eje de la cinta. En la cinta enrollada alrededor de un cilindro que aparece abajo a la izquierda, E, T y Z son nulos. Al dar dos vueltas completas a la parte superior del cilindro alrededor de su eje central en sentido antihorario, se forma la trenza helicoidal dextrorsa de abajo a la derecha, en la que E evidentemente no ha cambiado, pero (despreciando el efecto de los extremos) T es $+ 4\operatorname{sen}\alpha$ y por tanto Z vale $- 4\operatorname{sen}\alpha$. (En el primer caso se hizo que la torsión total T fuera cero, de modo que la configuración resultante no es sino un modelo simplificado del ADN superenrollado, que omite la torsión propia de las vueltas de la doble hélice.)



RELACION ENTRE TORSION Y ENZARCILLAMIENTO aplicada a un cable de teléfono. En su forma relajada (*arriba*) la torsión es pequeña y el enzarcillamiento grande; el eje del cable genera una hélice en el espacio tridimensional. Al estirar el cable (*en medio*) hasta que el eje queda casi recto (*abajo*) el enzarcillamiento es pequeño y la torsión grande. La torsión y el enzarcillamiento se compensan.



CONFIGURACIONES BASICAS DE LA DOBLE HELICE son la lineal (*izquierda*), la circular discontinua, y por tanto relajada (*centro*), y la circular cerrada (continua, sin roturas), y por tanto superenrollada (*derecha*). Las escalas de los tres dibujos son diferentes. El superenrollamiento de la molécula circular cerrada para adoptar la forma de una trenza puede explicarse por la relación entre enlace, torsión y enzarcillamiento. La molécula presenta un defecto de vueltas, por lo que su índice de enlace es menor que en moléculas relajadas del mismo tamaño; esta variación se compensa por tensión y enzarcillamiento, según la ecuación $Z = E - T$, y la molécula adopta la configuración de menor energía.

siones, cambiará el signo del índice de enlace, la torsión total y el enzarcillamiento. Uno de nosotros (White) y Thomas F. Banchoff, de la Universidad Brown, han encontrado otra operación matemática que cambia el signo de estas cantidades pero mantiene su valor absoluto: invertir la cinta respecto a una esfera. Así se explica que el enzarcillamiento de una cinta cuya curva eje esté contenida en la superficie de una esfera sea nulo: la operación descrita transforma la curva eje en sí misma.

Aunque la ecuación $Z = E - T$ demuestra que enlace y torsión son matemáticamente distintos, puede que no resulte evidente la diferencia física entre estas cantidades. Tal vez convenga considerar una cinta matemática liada sobre un cilindro de tal forma que su superficie se apoye siempre totalmente sobre el cilindro [*véase la ilustración de la página anterior*]. (Nótese que sólo una cinta matemática infinitamente delgada se puede adaptar así sobre un cilindro.) Llamaremos α a la inclinación de la hélice descrita por la cinta, o sea, al ángulo de la hélice con la horizontal, de modo que cuando α es pequeño la hélice es aplanada y, cuando α es grande, la hélice aparece estirada.

Supongamos que se enrolla la cinta alrededor del cilindro N veces y a continuación se unen sus extremos siguiendo el camino más corto posible. Ignorando el efecto de la unión de los extremos, se puede demostrar que E, el índice de enlace de la cinta, es igual a N, y que T, la torsión total, es $N\operatorname{sen}\alpha$. Por tanto, al estirar la hélice, aumentando el ángulo α, el número de vueltas de la hélice sigue siendo el mismo, y se mantiene igual el índice de enlace, pero aumenta la torsión, lo que pone claramente de manifiesto la diferencia entre uno y otro concepto.

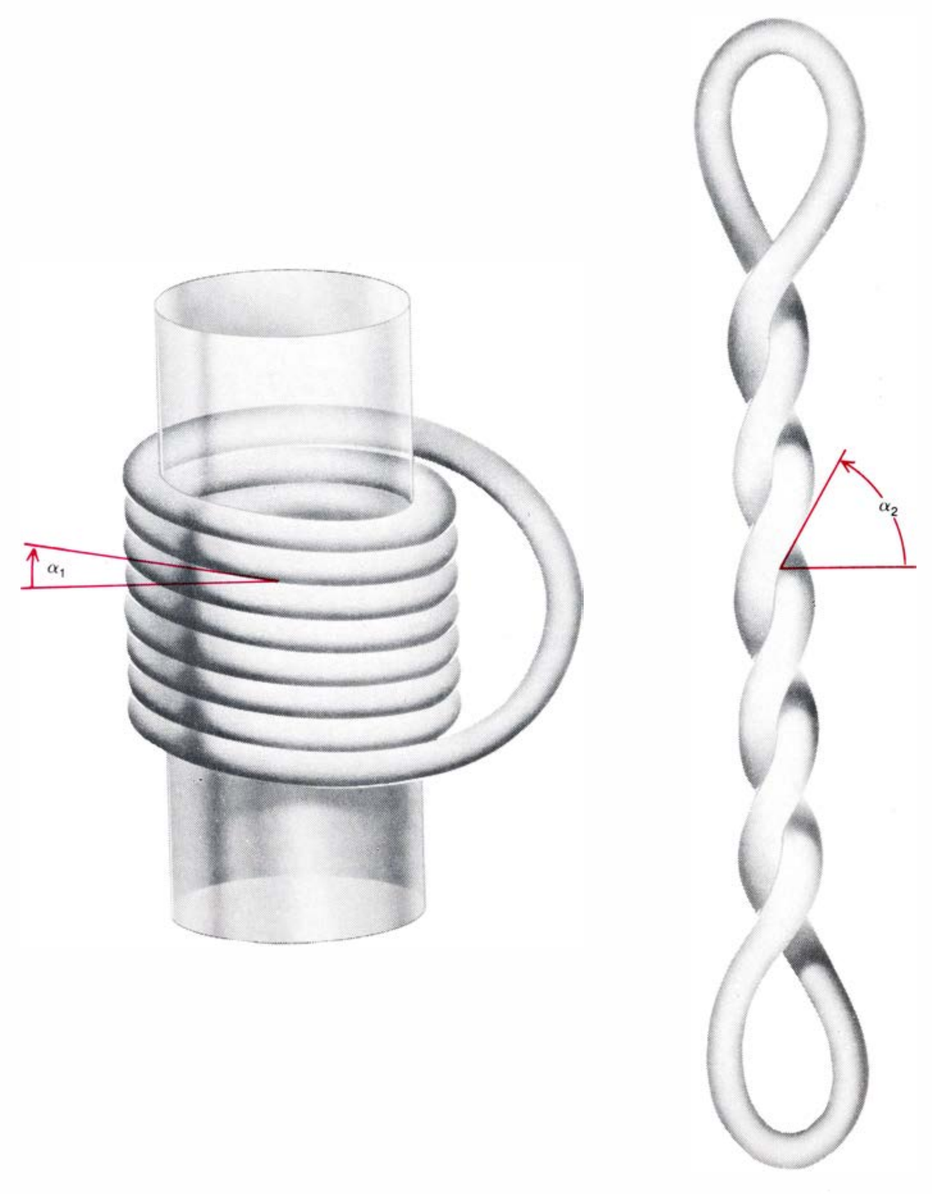
Además, puesto que el enzarcillamiento de una cinta es la diferencia entre su índice de enlace y su torsión total, el valor de Z será en este caso $N - N\operatorname{sen}\alpha$, o sea $N(1 - \operatorname{sen}\alpha)$. Como indica esta fórmula, si α, y por tanto la torsión, son pequeños, el enzarcillamiento es considerable, pero si α y la torsión son grandes, el enzarcillamiento es mínimo. Esta relación puede observarse fácilmente en un cable helicoidal de teléfono: en reposo está muy enzarcillado y la torsión es pequeña; al estirarlo aumenta la torsión y disminuye el enzarcillamiento

Apliquemos ahora estos resultados al ADN real, por ejemplo al ADN del virus polioma. Recordemos que este ADN se resuelve por sedimenta-

ción en tres componentes, I y II, circulares, y III, lineal. Se puede determinar experimentalmente que el índice medio de enlace de una población de moléculas circulares relajadas de este ADN es aproximadamente + 500. (El índice medio no es necesariamente entero, porque incluso en una preparación muy purificada no todas las moléculas están relajadas o superenrolladas en el mismo grado.) Por otra parte, en una población de moléculas circulares cerradas (las moléculas superenrolladas del componente I), el índice medio de enlace es aproximadamente + 475. Las moléculas circulares cerradas del ADN de polioma muestran un defecto de vueltas, como predijo Vinograd, y su índice de enlace tiene un déficit de alrededor de 25. Esta observación nos lleva a definir el superenrollamiento como la diferencia ΔE entre el índice de enlace de una molécula circular cerrada natural y el de la misma molécula en la forma circular cerrada relajada (en la que la energía de deformación es mínima y el enzarcillamiento nulo). Por ejemplo, los ADN de los virus polioma y SV40 tienen ΔE de aproximadamente $-$ 25. (Se supone que se mide el índice de enlace de moléculas relajadas superenrolladas en solución bajo condiciones normalizadas de temperatura, salinidad, etcétera.)

Podemos ahora comprender por qué un déficit en el índice de enlace de una molécula de ADN provoca el superenrollamiento. Una molécula lineal de ADN en solución adopta normalmente la estructura llamada configuración *B*, en que las bases quedan aproximadamente perpendiculares al eje de la hélice, con una distancia de 0,34 nanómetros entre bases sucesivas y unos 10 pares de bases por cada vuelta de la doble hélice. Esta es la configuración de energía mínima, pero si la molécula se curva o se retuerce, su energía aumenta. Si una molécula larga se circulariza, el diámetro del círculo es grande en comparación con la anchura de la doble hélice. La curvatura, por tanto, es pequeña y el aumento de energía escaso. Por eso, las moléculas circulares discontinuas, tales como el componente II del ADN de polioma, raramente abandonan la configuración *B*.

La situación cambia en moléculas circulares cerradas con un déficit en su índice de enlace. Para satisfacer la condición de que el valor de *E* fuera menor que en la molécula relajada (por ejemplo 475 en vez de 500), la doble hélice tendría que perder vueltas, y esto aumentaría considerablemente su energía de deformación. Superenrollándose, la molécula circular cerrada minimiza sus diferencias con la configuración *B*.



UN TUBO DE CAUCHO enrollado alrededor de un cilindro y con sus extremos unidos de modo que no haya torsiones (*izquierda*) salta a una forma en trenza de sentido opuesto al retirar el cilindro (*derecha*). Si la primera configuración tiene *N* vueltas sinistrorsas y una pequeña inclinación α_1 de la hélice respecto a la horizontal, la segunda configuración tendrá unas *N*/2 vueltas hacia arriba y *N*/2 hacia abajo, todas dextrorsas, y una inclinación α_2 grande. (Esta descripción es sólo aproximada.) El examen de las variaciones de índice de enlace, torsión total y enzarcillamiento que acompañan esta transformación explica por qué la mayoría de los ADN naturales, deficitarios en enlace, son superhélices dextrorsas.

Si la doble hélice tiene menos vueltas de las normales, el ADN puede reducir su energía de deformación enzarcillándose, como vimos al estudiar el modelo de cinta. La interconversión entre enzarcillamiento y torsión permite variar el grado de enzarcillamiento de forma que la torsión sea mínima, y mínima también la energía de deformación debida a la torsión. Por otra parte, el enzarcillamiento siempre introduce alguna curvatura, con lo que aumenta la energía de deformación debida a curvatura. Por tanto, el ADN con un defecto de vueltas adopta una configuración superenrollada que minimiza la torsión introduciendo la menor curvatura posible.

En consecuencia, si se provoca el enzarcillamiento del eje del ADN, la doble hélice responde mediante torsión o curvatura (o ambas a la vez) de la molécula. El superenrollamiento da lugar a superhélices de varias formas, desde el solenoide sinistrorso a la trenza dextrorsa, aunque en la naturaleza existe una preferencia por esta última (la trenza). Un ejemplo mecánico puede ilustrar mejor la compensación entre enzarcillamiento y torsión que da ventaja a la trenza. Consideremos un tubo de caucho, grueso y flexible, enrollado como un solenoide sinistrorso alrededor de un cilindro.

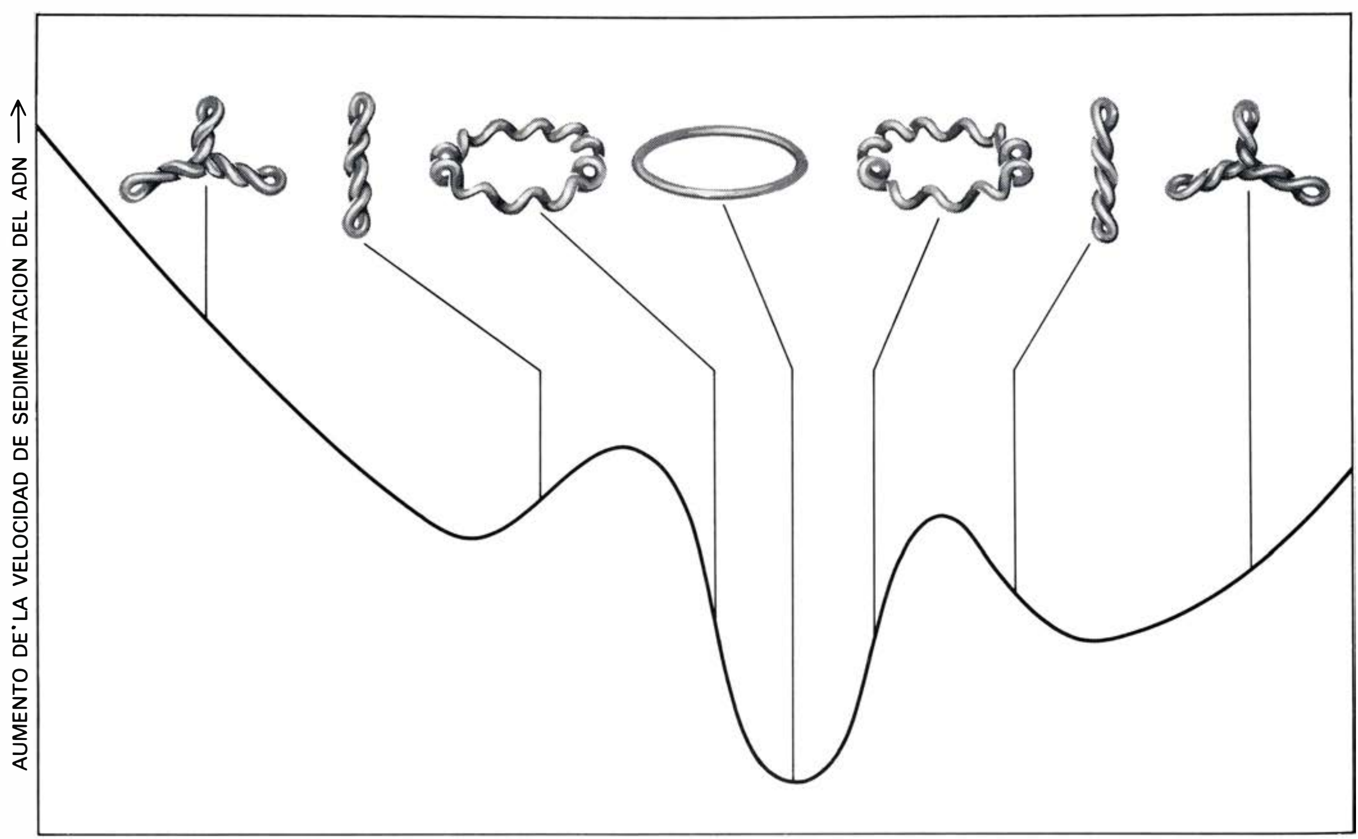
Supongamos que, al enrollar el tubo, se le ha dejado girar libremente alrededor de su propio eje, de modo que su energía de torsión sea nula. Se unen entonces sus extremos permanentemente, cuidando de evitar tensiones. Sin embargo, si quitamos el cilindro, el tubo adoptará la forma de una trenza dextrorsa [*véase la ilustración de la página anterior*]. ¿Por qué?

Sin entrar a fondo en los aspectos mecánicos de la situación, la comparación del índice de enlace, la torsión total y el enzarcillamiento de las dos configuraciones puede dar una idea general de por qué un solenoide sinistrorso, poco o nada retorcido pero muy enzarcillado, se transforma en una trenza dextrorsa. (Para dar sentido al índice de enlace en este caso, imagínense un par de líneas paralelas dibujadas en lados opuestos del tubo antes de enrollarlo.) Recuérdese que en la primera configuración anulamos deliberadamente la torsión total y por tanto, ignorando el efecto de la unión de los extremos, el índice de enlace debe ser igual al enzarcillamiento. En el ejemplo de la cinta helicoidal liada alrededor de un cilindro observamos que el valor del enzarcillamiento era igua a $-N_1(1 - \text{sen}\alpha_1)$, donde N_1 era el número de vueltas y α_1 la pendiente de la hélice. (El signo menos se debe a que la configuración es sinistrorsa.) Por tanto, el índice de enlace de la primera configuración será $-N_1(1 - \text{sen}\alpha_1)$. (Este número no será usualmente un entero, porque estamos despreciando el efecto de unión de los extremos.)

En cuanto a la segunda configuración del tubo de caucho, la trenza dextrorsa, llamaremos N_2 al número total de vueltas (en ambas direcciones) y α_2 a la inclinación de la superhélice. Olvidando de nuevo los extremos, se puede calcular que el enzarcillamiento Z vale $-N_2\text{sen}\alpha_2$. Nótese que el solenoide sinistrorso se transformó en trenza dextrorsa sin romper el tubo ni separar sus extremos. Por tanto, el índice de enlace E de la segunda configuración será el mismo de la primera, o sea $-N_1(1 - \text{sen}\alpha_1)$. Sustituyendo estos valores de Z y E en la ecuación $Z = E - T$, obtendremos $-N_2\text{sen}\alpha_2 = -N_1(1 - \text{sen}\alpha_1)$, es decir, $T = -N_1(1 - \text{sen}\alpha_1) + N_2\text{sen}\alpha_2$.

La inclinación α_1 de la primera configuración del tubo era pequeña, y la de la segunda, α_2, bastante grande. Podemos simplificar el cálculo introduciendo en la ecuación las aproximaciones $\text{sen}\alpha_1 = 0$ y $\text{sen}\alpha_2 = 1$, con lo que tenemos $T = N_2 - N_1$. Esta sencilla fórmula explica el cambio de configuración del tubo de caucho. En la fórmula aproximada, la torsión T es mínima cuando la segunda configuración tiene el mismo número de vueltas que la primera, pero en dirección opuesta. Además, la curvatura de la segunda configuración es claramente menor que la de la primera. Por tanto, si la torsión es pequeña, la energía de deformación de la trenza será mucho menor que la del solenoide y el tubo de goma adoptará aquella forma. (Para aplicar este argumento al ADN tendríamos que pintar la doble hélice sobre el tubo de caucho antes de enrollarlo, pero esto no afectaría al cálculo de la energía.)

El análisis detallado del superenrollamiento del ADN real requiere un conocimiento bastante preciso de todas sus constantes elásticas y hay que tener en cuenta, no sólo los efectos de los extremos, sino otros aspectos, como repulsiones electrostáticas y movimien-



LAS SUPERHELICES se relajan al añadir el colorante bromuro de etidio, de estructura molecular plana, a una solución de ADN superenrollado, como muestra esta gráfica del cambio de velocidad de sedimentación del ADN. La velocidad de sedimentación es aquella con que las moléculas de ADN se desplazan a través del solvente cuando se somete la solución al fuerte campo gravitatorio generado por una ultracentrífuga. Las moléculas más superenrolladas son más compactas y sedimentan más aprisa. Al aumentar la concentración de bromuro de etidio disminuye el enzarcillamiento del ADN, no porque varíe su índice de enlace (lo que exigiría cortar sus cadenas de nucleótidos), sino porque rebaja su torsión. Las moléculas del colorante se intercalan entre los pares de bases del ADN, y generan una disminución local de la torsión de la doble hélice. La gráfica muestra que con una concentración suficientemente alta del colorante las moléculas de ADN se relajan totalmente y que, a concentraciones mayores, el ADN se superenrolla en sentido opuesto.

tos térmicos. El argumento precedente muestra, sin embargo, con bastante claridad la causa del superenrollamiento de una molécula circular cerrada de ADN con un déficit en el índice de enlace (o, en algunos casos, con un exceso). En la naturaleza, el superenrollamiento de la mayoría de las moléculas de ADN circular cerrado es negativo, es decir, debido a un déficit en el índice de enlace. El ejemplo del tubo de caucho explica por qué cabe esperar que tales moléculas adopten la configuración de una trenza que será, sorprendentemente, igual que la de la doble hélice, es decir, dextrorsa.

Las cadenas de la doble hélice son lo bastante flexibles para que, si se corta una de ellas, la otra pueda girar a su alrededor y relajar la molécula. Las moléculas continuas, sin incisiones, no pueden perder su superenrollamiento de esta manera, porque una de las cadenas habría de atravesar a la otra, lo que es físicamente imposible. Dada una molécula superenrollada natural de ADN, con índice de enlace deficitario, ¿sería posible disminuir su enzarcillamiento sin romper ninguna de las cadenas? Si no se rompen cadenas, el índice de enlace no cambia y como el enzarcillamiento es, en este caso, negativo, el único modo de disminuir el enzarcillamiento sería disminuir la torsión total.

Esto se consigue añadiendo bromuro de etidio, un colorante de estructura molecular plana, a la solución de ADN. Leonard S. Lerman, de la Universidad estatal de Nueva York en Albany, descubrió que las moléculas de este tipo se intercalaban entre los pares de bases de la doble hélice, causando una disminución local de la torsión de las dos cadenas. De hecho, con una concentración suficientemente alta de bromuro de etidio se consigue relajar completamente la molécula.

Más exactamente, esta concentración crítica disminuirá la torsión total de la molécula. Por ejemplo, en presencia del colorante, una molécula lineal de ADN del virus polioma tendrá una torsión total de 475 en vez de 500. Por tanto, una molécula circular cerrada de ese ADN con un índice de enlace de 475 tendrá, en presencia de esa concentración crítica del colorante, una torsión total igual a su índice de enlace, con lo que su enzarcillamiento será nulo. La molécula adoptará la configuración relajada. Si la concentración de bromuro de etidio es mayor que la crítica, su torsión total será aun menor y la molécula adoptará un superenrollamiento de sentido opuesto.

Estos efectos se pueden observar directamente estudiando la velocidad de sedimentación del ADN, como demostraron por primera vez Lionel V. Crawford, del Instituto Imperial británico para Investigación del Cáncer, y Michael J. Waring, de la Universidad de Cambridge e, independientemente, Vinograd y uno de nosotros (Bauer). La adición de bromuro de etidio a una solución de moléculas circulares cerradas de ADN de polioma disminuye su enzarcillamiento, por lo que se vuelven menos compactas y sedimentan más despacio. A cierta concentración de bromuro de etidio, la velocidad de sedimentación es mínima; a concentraciones mayores, las moléculas se superenrollan en sentido opuesto, devienen más compactas y la velocidad de sedimentación vuelve a aumentar.

Para crear moléculas de ADN con distintos grados de enzarcillamiento convendría cortar una de las cadenas de la doble hélice, relajar la molécula dejándole dar unas pocas vueltas y cerrar la incisión. Sorprendentemente, se han identificado enzimas capaces de hacerlo. El primero, la llamada proteína ω, fue descubierto en *Escherichia coli* por James C. Wang, de la Universidad de Harvard, en 1971. Se han obtenido enzimas similares de varias fuentes, entre ellas otras bacterias, el virus de la vacuna y los núcleos y mitocondrias de los animales. Estos enzimas, los denominados topoisomerasas, no suelen necesitar una fuente de energía para funcionar. Reducen siempre el superenrollamiento del ADN y con ello disminuyen su energía.

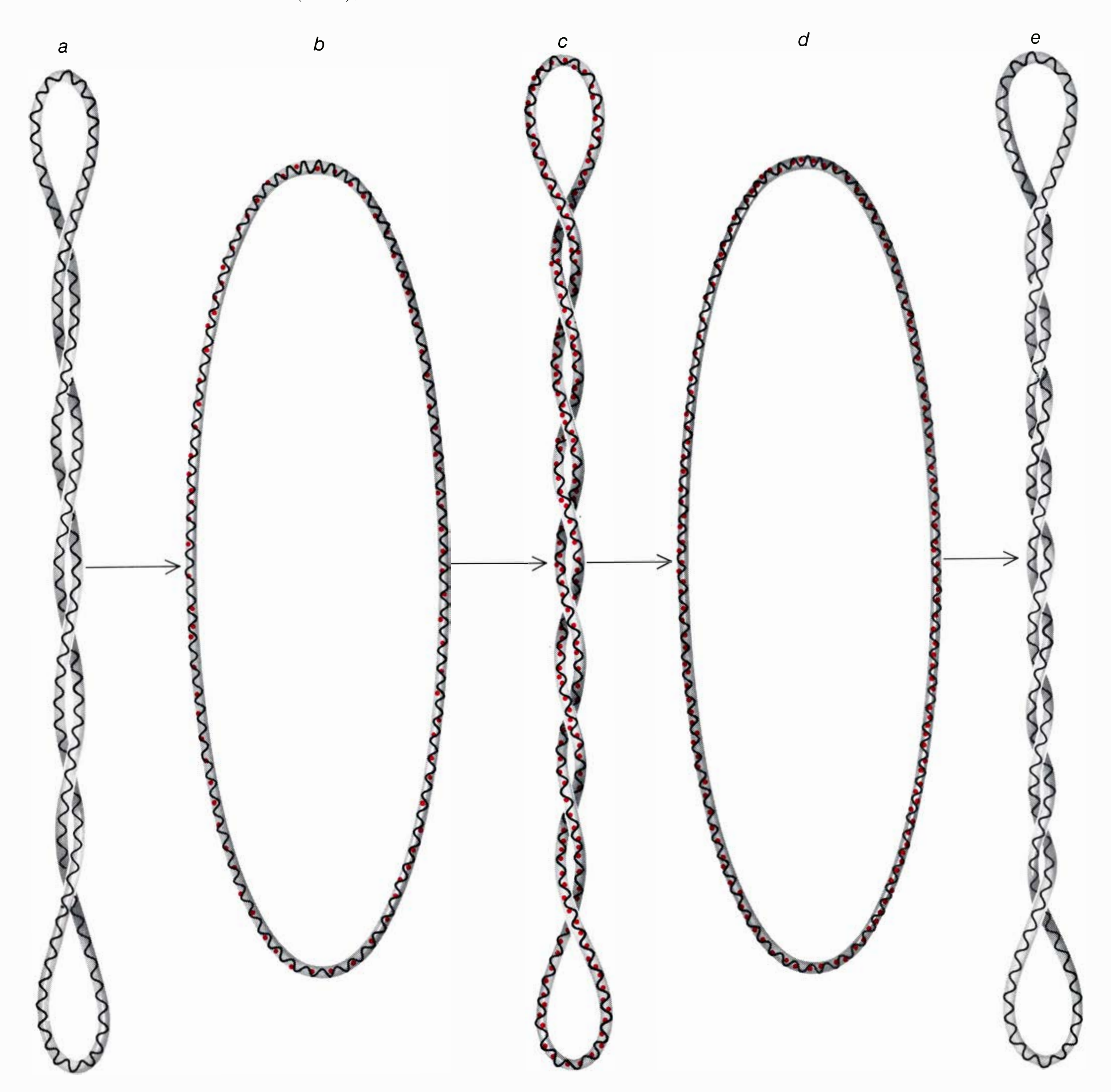
Las topoisomerasas han resultado de gran utilidad en el estudio de la química física del ADN. Para empezar, la adición de topoisomerasa a una solución de ADN superenrollado disminuye gradualmente el superenrollamiento, hasta que todas las moléculas están aproximadamente en el estado relajado; pero estos enzimas tienen aplicaciones mucho más ingeniosas. Por ejemplo, supongamos que se añade una alta concentración de bromuro de etidio a ADN circular cerrado, superenrollándolo en sentido opuesto al usual. Si se añade además una topoisomerasa, las moléculas se relajarán, pero dada la gran cantidad de bromuro de etidio intercalado tendrán bastante menos vueltas de lo que es natural. Si primero se elimina el enzima y luego el colorante, las moléculas compensarán su bajísimo índice de enlace superenrollándose aún más en la dirección usual. De esta manera se obtienen moléculas hiperenrolladas, que tienen muchas

más vueltas de superhélice de lo corriente. Este tipo de almacenamiento y transferencia de energía puede ser crucial para la función celular del superenrollamiento. Incluso se puede imaginar un motor químico que, pasando por un ciclo similar, convirtiera energía química en trabajo.

Es también interesante señalar que Martin Gellart y sus colegas, de los National Institutes of Health, encontraron un enzima de *E. coli* que, en presencia de una fuente de energía como el trifosfato de adenosina (ATP), disminuía el índice de enlace del ADN circular relajado, aumentando su enzarcillamiento y superenrollándolo. Los enzimas de este tipo, las girasas, efectúan esencialmente lo contrario que las topoisomerasas. Hasta ahora se han encontrado girasas en varios microorganismos, pero no en organismos superiores.

Nicholas R. Cozzarelli y sus colegas, de la Universidad de Chicago, han descubierto recientemente que una girasa parece actuar acercando dos segmentos de la molécula de ADN, cortando las dos cadenas de uno de los segmentos, pasando el segmento intacto a través de la brecha abierta y reuniendo las cadenas cortadas. Es fácil calcular que este proceso altera el índice de enlace del ADN en dos unidades y no en una unidad. Eso es precisamente lo observado experimentalmente.

¿Cómo se mide el superenrollamiento? Un método experimental se basa en un descubrimiento de Walter Keller, de la Universidad de



ADN HIPERENROLLADO, con más vueltas de superhélice de lo normal, se puede obtener con bromuro de etidio (*puntos de color*) y una topoisomerasa. Esta reduce el superenrollamiento del ADN cortando una de sus cadenas, relajando la molécula y reparando el corte. Si se añade a la solución suficiente bromuro de etidio (*a*) para que el ADN se relaje completamente (*b*) y después se superenrolle en sentido opuesto (*c*), la topoisomerasa volverá a relajar el ADN (*d*). El ADN tendrá intercaladas muchas moléculas de colorante, por lo que en sus cadenas se apreciará un defecto de vueltas mayor de lo usual. En consecuencia si se retira primero el enzima y luego el colorante, el ADN se superenrollará más de lo que suele hacer, en el sentido original (*e*).

Heidelberg, en 1974. Keller observó que, al someter a electroforesis en gel de agarosa moléculas superenrolladas con índices de enlace tan sólo ligeramente distintos, no se desplazaban a la misma velocidad. Las moléculas más compactas iban más deprisa, y así aparecía en el gel una serie regular de bandas. Además, como más tarde demostró Vinograd, si se cortaba una de las bandas y se repetía el experimento con sus moléculas, se observaba de nuevo una sola banda exactamente en el mismo lugar que antes. Incluso, como descubrieron posteriormente Wang y uno de nosotros (Bauer), si se calentaba la banda para separar las bases complementarias de las dos cadenas y se enfriaba lentamente, con lo que las cadenas se unían de nuevo, aún se obtenía una sola banda, y en el mismo lugar.

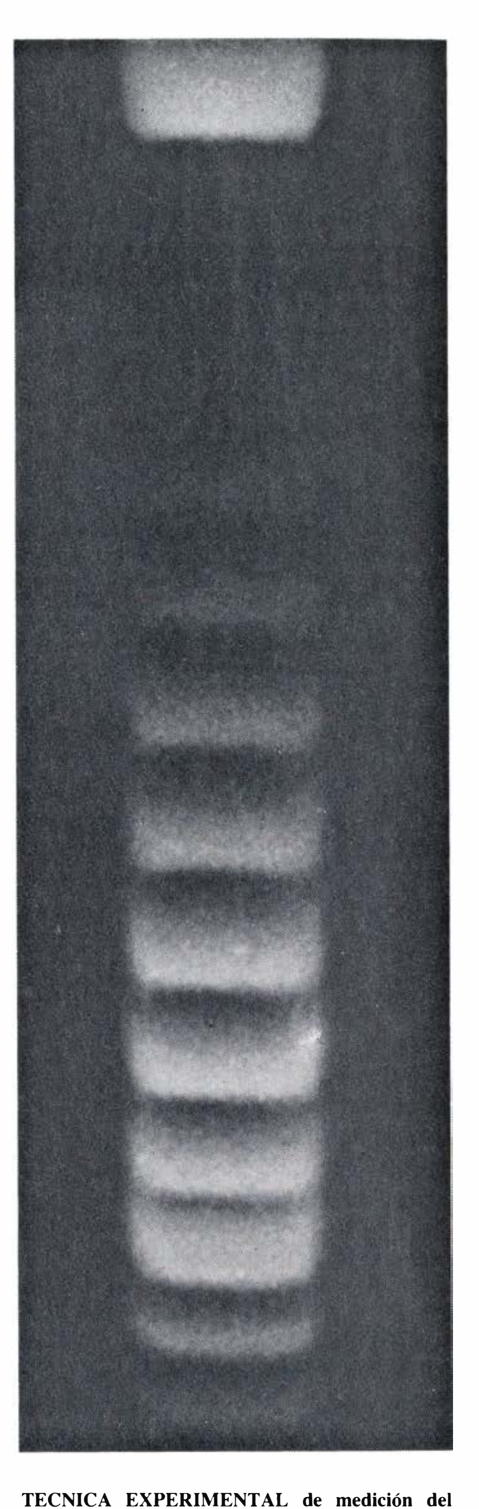
Las moléculas de bandas adyacentes difieren por tanto en una cantidad inalterada por el calentamiento y el enfriamiento, cantidad que debe ser un valor discreto, dada la regularidad de las bandas. Esta cantidad sólo puede ser el índice de enlace. La única interpretación razonable es que los índices de enlace de las moléculas de bandas adyacentes en el gel difieren en una unidad. Por tanto, ΔE, la diferencia en el índice de enlace responsable del superenrollamiento, se puede determinar directamente contando las bandas obtenidas por electroforesis de una población de ADN que incluya moléculas en todos los estados, desde relajadas a superenrolladas. Este valor experimental de ΔE puede ser útil, por ejemplo, para medir la distorsión de la doble hélice producida por el bromuro de etidio. Resulta que cada molécula intercalada disminuye la torsión en unos 26 grados. Este dato permite utilizar el bromuro de etidio para medir directamente el valor de ΔE de cualquier molécula superenrollada. También se puede observar claramente en geles el sorprendente efecto de la girasa sobre el índice de enlace.

Uno de los resultados más importantes surgidos hasta ahora del estudio del ADN superenrollado es la prueba directa de que las dos cadenas dan efectivamente una vuelta aproximadamente cada 10 pares de bases. Los experimentos más elegantes en este sentido fueron hechos por Wang. Comparó moléculas circulares de ADN construidas por métodos de la nueva ingeniería genética, cada una de unos 5000 pares de bases. no se conocía la longitud exacta de estas moléculas, pero sí la diferencia entre cada dos de ellas. Algunas diferían sólo en un par de bases y otras hasta en 400.

Las diferencias de superenrollamiento entre estas distintas moléculas, observadas tras relajación por topoisomerasas y electroforesis, permitieron a Wang calcular que cada vuelta de la doble hélice tenía 10,4 ± 0,1 pares de bases. Su experimento fue una hermosa demostración de la precisión y refinamiento de las técnicas actuales de la biología molecular. Además, se han atribuido recientemente al ADN estructuras en las que las dos cadenas no dan vueltas una alrededor de otra para formar la doble hélice, sino que corren contiguas casi siempre, con pocas vueltas de hélice. Los resultados de Wang y los de otros experimentos relacionados demuestran que estas estructuras son incorrectas. Este argumento topológico elimina todos los modelos de cadenas vecinas no enrolladas, cualesquiera que sean sus detalles moleculares.

¿Cómo se origina el superenrollamiento? En los virus polioma y SV40, el ADN se halla superenrollado porque no suele estar desnudo en el núcleo de la célula. Durante la replicación, la doble hélice se enrolla sobre nucleosomas (gránulos de proteína compuestos por ocho moléculas de histona y una o dos moléculas asociadas). Al desprenderse de este soporte, la molécula de ADN se superenrolla, de modo parecido al tubo de caucho de que hablamos antes. El bajo índice de enlace del ADN natural implica que el solenoide formado al enrollarse el ADN sobre los nucleosomas sea sinistrorso. La imprecisión de los datos disponibles y la incertidumbre de su interpretación dificultan el cálculo del número exacto de vueltas por nucleosoma, pero es probable que se hallen entre una y dos. En todos los organismos superiores, el ADN está asociado a nucleosomas; en los organismos inferiores los orígenes del superenrollamiento están menos claros.

Queda mucho por hacer hasta alcanzar una comprensión total de las consecuencias del superenrollamiento del ADN. En la actualidad se investigan intensamente los enzimas relacionados con el superenrollamiento (topoisomerasas y girasas) y el papel del superenrollamiento en la replicación. Está claro, sin embargo, que el conocimiento actual del superenrollamiento, tanto teórico como experimental, es una buena base para la investigación de este aspecto sorprendente de la doble hélice.



TECNICA EXPERIMENTAL de medición del superenrollamiento. Se hace correr una población de moléculas circulares cerradas de ADN por electroforesis a través de un gel de agarosa, de arriba abajo en este caso. Como las moléculas más superenrolladas son más compactas, se desplazan por el gel a mayor velocidad. Se forma así una serie regular de bandas, en que las moléculas de cada banda tienen el mismo índice de enlace y los índices de bandas adyacentes difieren en una unidad. (En este caso el ADN se tiñó con un colorante fluorescente, por lo que brillan más las bandas que tienen más moléculas.) Si la población contiene moléculas con varios grados de superenrollamiento, como en este caso, el déficit de índice de enlace, responsable del grado de superenrollamiento, se puede medir directamente contando las bandas. La banda superior contiene ADN relajado. Fotografía proporcionada por J. Vinograd.

La hélice

Las hélices marinas y las de avión tienen aspectos diferentes, pero ambas se diseñan de acuerdo con la teoría circulatoria de la sustentación. La teoría es la tradicional, desarrollada para determinar la sustentación de un perfil

E. Eugene Larrabee

En 1845, en los comienzos de la historia de la navegación a vapor, la eficiencia de la hélice quedó establecida cuando el buque de vapor *Rattler,* equipado con ella, arrastró de popa al buque de vapor *Alecto,* equipado con rueda de paletas, a una velocidad aproximada de tres nudos, en una competición entre ambos organizada por el almirantazgo británico. Ambos buques, muy semejantes, estaban accionados por motores de vapor de baja velocidad, de aproximadamente 200 caballos de vapor (CV). El ensayo fue significativo en cuanto demostró la validez de la propulsión por hélice frente al concepto, intuitivamente más comprensible, de la rueda de paletas.

La hélice ha seguido siendo el medio más utilizado para accionar vehículos náuticos de todos los tipos y tamaños. Las excepciones más importantes son algunos vehículos especiales, tales como los de poco calado o hidrodeslizadores. Cuando se trata de propulsar un avión provisto de motor alternativo, no existe otra opción que la hélice. La hélice se ha mantenido también en los aviones a turbohélice accionados por turbinas de gas. La mayoría de los aviones de líneas comerciales, sin embargo, se propulsan mediante motores turbofan, en que la hélice se conoce bajo otro nombre. En un motor turbofan, los gases que salen a alta velocidad de la cámara de combustión fluyen a través de la turbina de potencia, que acciona a un "fan", o compresor axial, en la parte anterior del motor. El fan es en realidad una hélice de muchas palas, diseñada para operar dentro de la carena del motor. El fan proporciona el aire para la combustión y crea un flujo de aire de baja velocidad y alto volumen que no pasa por la cámara de combustión y que se mezcla con los gases calientes de la salida de la turbina de potencia para propulsar el avión. En el avión de turbohélice, la turbina de potencia acciona una hélice exteriormente visible que proporciona la mayor parte del empuje. Como veremos, el turbohélice, gracias a su buen rendimiento, puede estar a punto de ganar aceptación nuevamente.

Cualquiera que haya observado atentamente la hélice de una lancha motora y una hélice típica de avión, se habrá dado cuenta de que las palas de la hélice marina cubren una proporción mucho mayor de la superficie total del disco, área barrida por las palas de la hélice, de la que cubren las palas de la hélice de un avión. El ingeniero aeronáutico diría que la hélice marina tiene mayor "solidez". Se puede observar un contraste extremo si se compara la hélice marina de un petrolero gigante con la hélice del *Gossamer Albatross,* el avión ultraligero que Bryan Allen propulsó a pedal a través del Canal de la Mancha el 12 de junio del pasado año. Las hélices de los grandes petroleros tienen 5 o 6 palas, un diámetro de hasta 9 metros, giran a unas 95 revoluciones por minuto y pueden absorber 45.000 CV para propulsar un barco de más de 500.000 toneladas a 16 nudos (8,23 metros por segundo). Las hélices gemelas del *Queen Elizabeth 2,* que se desplaza a 31 nudos, absorben cada una 45.000 CV, son apreciablemente más pequeñas, giran mucho más deprisa y tienen una solidez aún mayor que la de las hélices de los mayores petroleros.

La hélice del *Gossamer Albatross* posee dos palas esbeltas y un diámetro del 70 por ciento del de las hélices del *Queen Elizabeth 2;* cuando la hélice de 4,1 metros gira a 130 revoluciones por minuto, absorbe 0,25 CV para propulsar un avión con un peso en vuelo de 96 kilogramos a 5,4 metros por segundo. Las hélices marinas y las hélices de avión son, ambas, el resultado de análisis similares en ordenador, que tienen en cuenta la densidad del fluido, la interacción de los campos de corriente, la carga hidrodinámica o aerodinámica sobre las palas y finalmente las limitaciones prácticas impuestas por el vehículo y su modo de operación. ¿Por qué son las geometrías de las hélices tan diferentes?

No es porque el aire sea un gas y el agua un líquido. Curiosamente, la hélice del *Gossamer Albatross* se aproxima mucho más que la de barco a la forma ideal de una hélice de máxima eficiencia adaptada a la operación en un fluido casi incompresible de baja viscosidad, esto es, un fluido tal como el agua. Su geometría se adapta a las consideraciones teóricas establecidas, en 1919, por Albert Betz y Ludwig Prandtl, del Instituto Kaiser Wilhelm para Investigación en Dinámica de Fluidos de Göttingen, y refinadas por un científico británico visitante, Sidney Goldstein, en 1929. Sus ideas surgieron de la revolución en hidrodinámica teórica, a comienzos de siglo, cuando Prandtl inventó la teoría de la capa límite, que explicaba la resistencia de cuerpos de forma aerodinámica, y cuando W. M. Kutta, de Alemania, y Nikolai Joukowsky, de Rusia, idearon, independientemente el uno del otro, la teoría circulatoria de la sustentación, que explicaba la sustentación, tanto de las alas como de las palas de hélice.

Hasta entonces, la hidrodinámica teórica, surgida del estudio de la resistencia al avance de cascos de buque en movimiento, había sido en su mayor parte una disciplina académica que se ocupaba de los campos matemáticos de corriente que podían establecerse mediante diversas combinaciones de campos. Aunque podía hacerse que las líneas de corriente de alguno de los campos matemáticos de corriente combinados se pareciesen a la corriente alrededor de un buque, Jean le Rond d'Alembert había demostrado en 1742 que la resistencia teórica de tal corriente fluida sobre la forma matemática del buque era exactamente igual a cero. Es más, si la forma del casco se giraba matemáticamente un cierto ángu-

HELICE DEL "QUEEN ELIZABETH 2" que con su gemela absorben en conjunto 110.000 CV en la propulsión del buque de 67.000 toneladas brutas, a una velocidad máxima de 31 nudos. La hélice, que mide 5,8 metros de diámetro, genera un empuje máximo cuando gira a 174 revoluciones por minuto. El diámetro de las hélices náuticas está fuertemente limitado, pues no debe proyectarse por debajo de la quilla o más allá del costado del buque. A fin de alcanzar el empuje necesario dentro del límite del diámetro, la superficie total de las palas supone un 89,5 por ciento de la superficie de un disco del mismo diámetro. En la terminología marina, la relación superficie de pala a superficie de disco se denomina razón de superficie extendida; en aeronáutica, esa misma relación se conoce por solidez. La hélice del *Queen Elizabeth 2* fue proyectada y construida por SMM Propeller Ltd. Liverpool.



HELICE DEL "GOSSAMER ALBATROSS", avión ultraligero que Bryan Allen propulsó a pedal a través del Canal de la Mancha el 12 de junio de 1979. Esta hélice guarda sólo una remota semejanza con la del *Queen Elizabeth 2*, que aparece en la figura superior. Ambas, no obstante, actúan "empujando" y se cuentan entre las de mayor eficiencia jamás diseñadas. La hélice del *Gossamer Albatross* mide 4,1 metros de diámetro, gira a 95 revoluciones por minuto y absorbe 0,8 caballos de vapor (CV) en el despegue y 0,25 CV a velocidad de crucero. La hélice se diseñó mediante ordenador con el algoritmo ideado por el autor, de forma que tuviera el mínimo de energía en la estela del torbellino y la menor pérdida por fricción en el vuelo de ascenso suave. El *Gossamer Albatross* fue diseñado por la Aero Vironment Incorporated de Pasadena, California. Aparece aquí en el Johnson Space Flight Center.

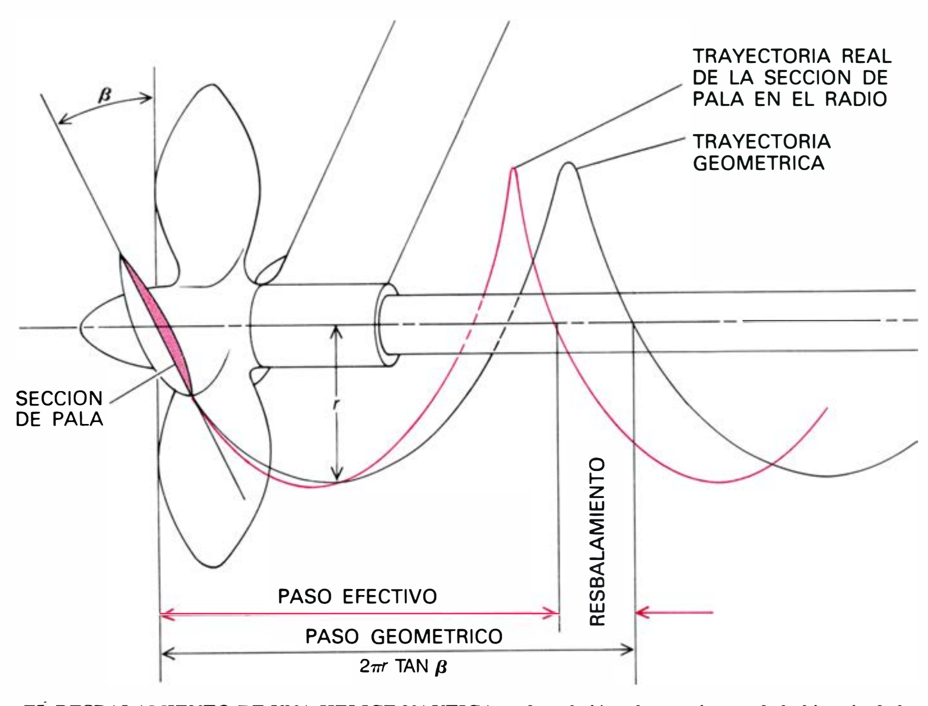
lo respecto del campo de corriente, la fuerza hidrodinámica perpendicular a la dirección de desplazamiento del casco era también nula. Como los cascos reales de buques no ofrecían resistencia nula a las corrientes reales, estos resultados recibieron el nombre de paradoja de d'Alembert.

En ausencia de una teoría útil, la arquitectura naval del siglo XIX se transformó en una ciencia experimental y las hélices marinas se diseñaron intuitivamente. La hélice se consideró semejante a un tornillo que avanzaba al girar en un orificio roscado. A diferencia del metal, sin embargo, el agua fluye bajo el empuje de la hélice. En consecuencia, el "paso efectivo", o distancia avanzada en una revolución, es menor que el "paso geométrico" de la hélice y difiere de él en una magnitud llamada deslizamiento. En 1875, el ingeniero escocés, William J.M. Rankine, desarrolló una teoría de la hélice llamada del "disco impulsor", que fue mejorada más tarde por William Froude, en la cual el deslizamiento podía explicarse como consecuencia de un simple incremento en la cantidad de movimiento del fluido en la estela de la hélice, un incremento proporcional al empuje de la misma.

En esta teoría, el deslizamiento es la mitad del incremento final en la velocidad de la estela (que depende de la tracción por unidad de área del disco y de la velocidad del buque) multiplicado por el tiempo necesario para una revolución de la hélice. El paso geométrico debe incrementarse en consecuencia, cuando uno trata de conseguir el paso efectivo (o deseado), que es la relación entre la velocidad del buque en pies por minuto y la velocidad angular del árbol en revoluciones por minuto. Aunque todo esto es casi experimentalmente cierto, la teoría del disco impulsor no proporcionaba información de por qué motivo las palas desarrollaban cargas y pares hidrodinámicos en primer lugar. Dio sin embargo un precedente para la idea de velocidad inducida en el disco, que fue identificada más tarde con la velocidad inducida de la teoría del ala, y asoció un resbalamiento excesivo (y una carga discal elevada) con una eficiencia reducida.

La explicación de las cargas de tracción y de par no pudo efectuarse hasta el desarrollo de la teoría turbillonaria, o circulatoria, de la sustentación de Kutta y Joukowsky en el período entre 1902 y 1911. En la versión de Joukowsky de la teoría, la corriente ideal de un fluido no viscoso alrededor de un cilindro circular se transforma en el movimiento alrededor de un perfil con borde de salida afilado. Se añade circulación (o vorticidad) a la corriente alrededor del cilindro introduciendo un torbellino lineal "ligado" de intensidad arbitraria situado en el centro del cilindro. La intensidad del torbellino se ajusta matemáticamente hasta que la corriente alrededor del cilindro presenta un punto de remanso en un punto singular que coincide en la transformación precisamente con el borde de salida del perfil.

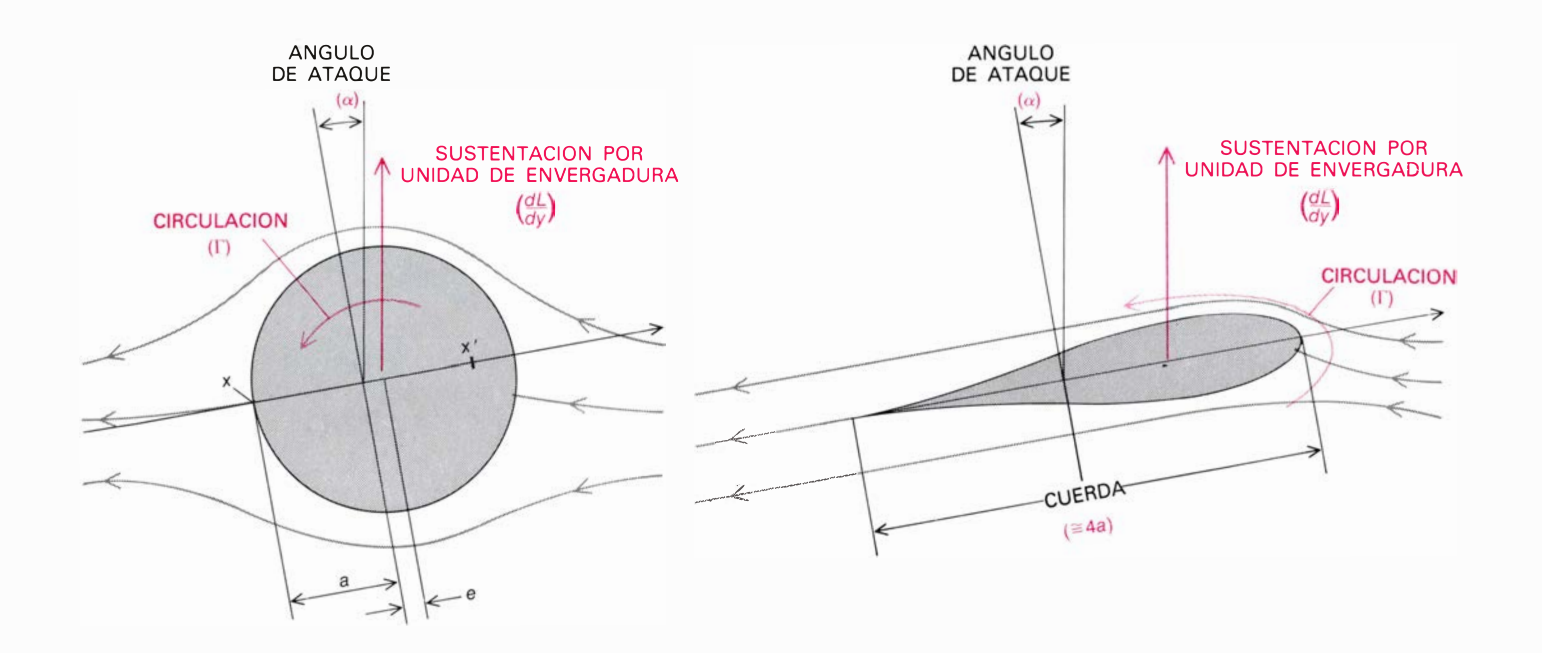
La hipotética circulación de un fluido no viscoso alrededor de un cilindro de longitud infinita se parece superficial-



EL RESBALAMIENTO DE UNA HELICE NAUTICA se descubrió en los comienzos de la historia de la hélice marina. La distancia recorrida verdaderamente por el barco durante una revolución de la hélice, el "paso efectivo", es menor que el paso geométrico determinado por el ángulo de pala de la hélice. Rankine y Froude explicaron que el deslizamiento correspondería al incremento de cantidad de movimiento axial impartido a la estela por la hélice. Un deslizamiento excesivo conducirá a un rendimiento bajo.



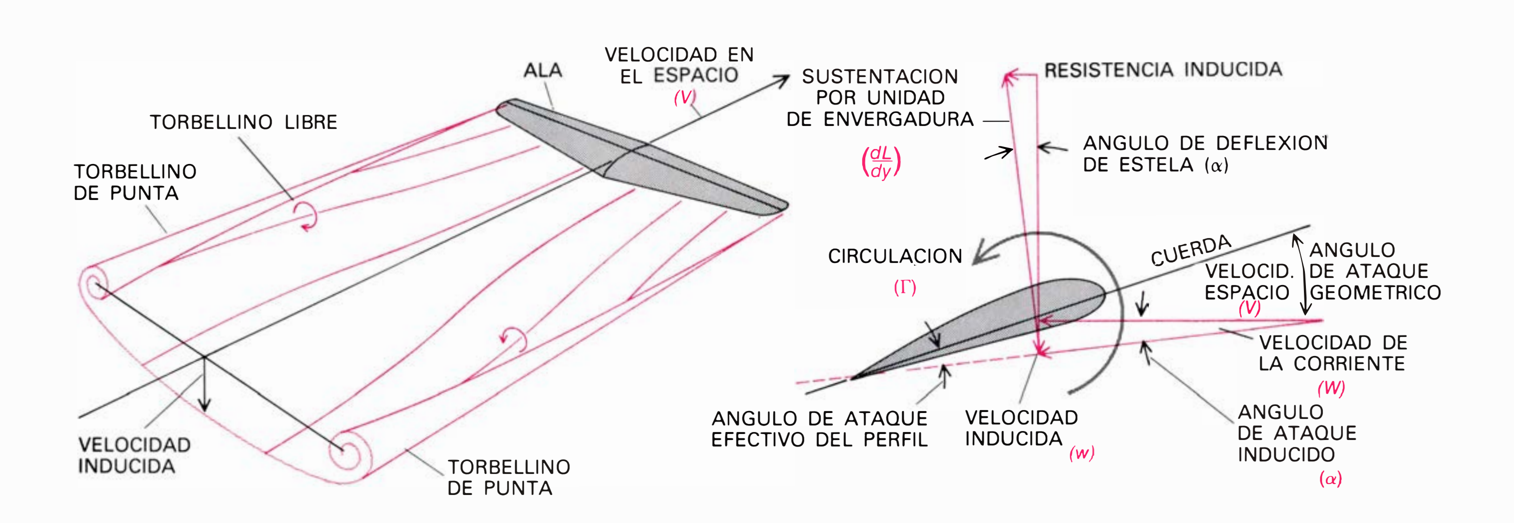
UNA PELOTA DE TENIS que avanza y gira sobre sí misma en sentido indicado en la figura se comporta en forma análoga a un perfil aerodinámico. Distorsiona la corriente de tal manera que da lugar a sustentación aerodinámica. La primera explicación de la trayectoria de una pelota de tenis que avanza y gira de la forma indicada se debe a Lord Rayleigh. A causa del giro de la pelota, el aire que fluye por su parte superior se acelera y el que fluye por la parte inferior se retrasa. Este efecto se incrementa debido a la superficie rugosa de la pelota. De acuerdo con el teorema de Bernoulli, sobre la conservación de la energía, la presión en la parte acelerada del fluido por encima de la pelota debe disminuir con respecto a la presión en la parte decelerada de fluido por debajo de ella. El desequilibrio levanta la pelota.



SUSTENTACION POR UNIDAD DE ENVERGADURA

$$= \left(\frac{dL}{dy}\right) = \rho V \Gamma$$

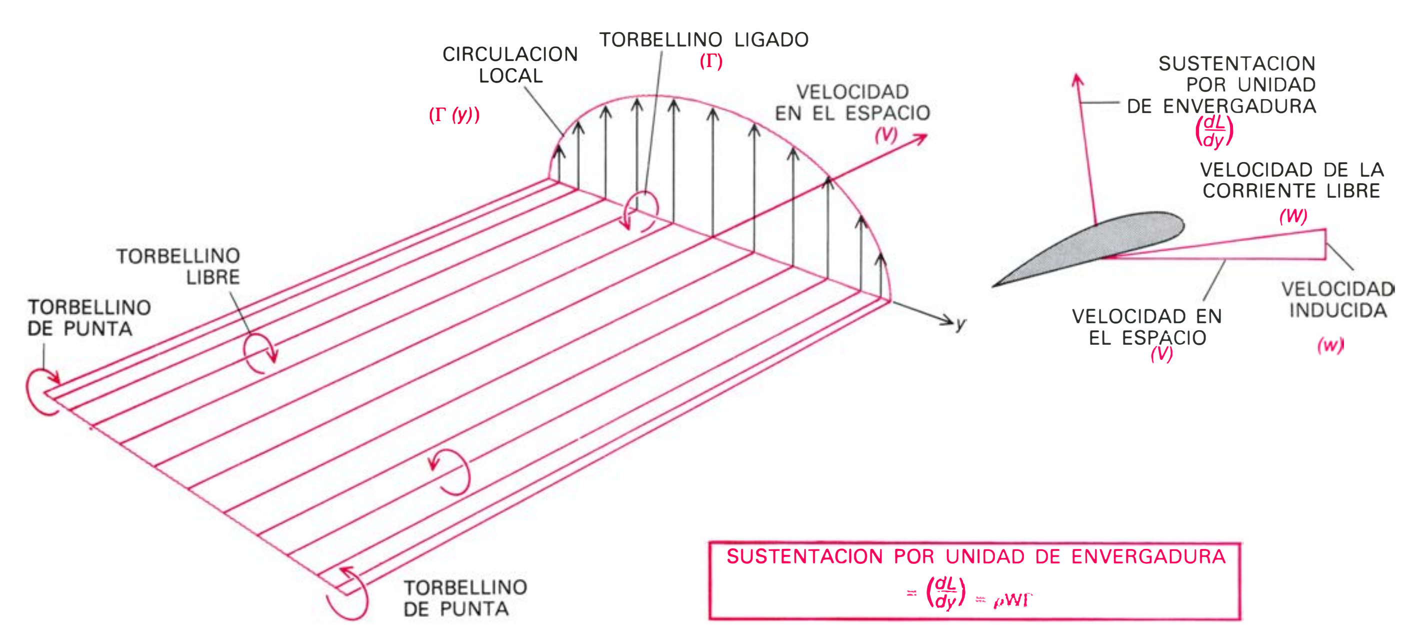
TEORIA CIRCULATORIA DE LA SUSTENTACION, propuesta independientemente entre 1902 y 1911 por W. M. Kutta, de Alemania, y Nikolai Joukowsky, de Rusia. En la formulación de Joukowsky, que se representa en la figura, la corriente de un fluido no viscoso alrededor de un cilindro de longitud infinita se transforma matemáticamente en la corriente alrededor de un perfil. La corriente puede ser simétrica, como en este ejemplo. Joukowsky añadió a la corriente no viscosa la debida a un torbellino ideal situado en el centro del cilindro. La intensidad de este torbellino "ligado" a circulación, (Γ), se elige de forma que dé lugar a un punto de remanso en la corriente situado en el punto singular posterior *x* sobre el cilindro. Cuando esta circulación se transforma sobre el perfil, alinea la corriente con el borde de salida afilado del mismo. Bajo estas condiciones, la circulación genera sustentación tanto sobre el cilindro como sobre el perfil. La sustentación por unidad de envergadura es igual a ϱ *V* Γ, donde ϱ es la densidad del fluido (en este caso el fluido es desde luego aire) y *V* es la velocidad de la corriente remota, o no perturbada. El espesor del perfil se controla mediante el valor que se asigne a la magnitud del desplazamiento *e*. El ángulo de ataque del perfil, α, corresponde a la desviación de los dos puntos singulares de transformación, designados por *x* y *x'*, con respecto a la dirección de la corriente remota. Esta teoría se aplicaba a las alas y a las palas de hélice.



SUSTENTACION POR UNIDAD DE ENVERGADURA

$$= \left(\frac{dL}{dy}\right) = \rho W \Gamma$$

CAPA DE TORBELLINOS LIBRES, concepto introducido por Ludwig Prandtl para justificar la corriente alrededor de las alas, donde la sustentación y la vorticidad ligada deben disminuir a cero en las puntas. La capa de torbellinos (*izquierda*) está formada por filamentos o líneas infinitesimales de torbellinos, en buena alineación con la dirección de vuelo y con intensidad creciente hacia los bordes marginales del ala, donde se arrollan para formar torbellinos de punta. El campo colectivo de velocidades de los torbellinos libres imparte una velocidad inducida a la capa de torbellinos como parte del proceso de arrollamiento. Tal como se muestra a la derecha, la velocidad inducida (*w*) se suma vectorialmente a la velocidad del vuelo en el espacio (*V*) para dar lugar a una velocidad de la corriente local (*W*) que forma con la velocidad de vuelo el ángulo de ataque inducido (α). Para producir una sustentación determinada, el ángulo de ataque geométrico de la sección de ala debe exceder al ángulo de ataque efectivo de la misma en un valor igual al ángulo de ataque inducido. El vector sustentación (*dL/dy*) gira hacia atrás un ángulo igual al inducido, creando una resistencia inducida.



DISTRIBUCION OPTIMA DE LA SUSTENTACION, correspondiente a la mínima resistencia inducida. Dicha distribución se consigue cuando la sustentación y la circulación "ligada" varían elípticamente de punta a punta del ala. Con esa distribución, el campo de velocidades inducidas por el sistema de torbellinos libres desplaza a la capa de torbellinos hacia abajo con velocidad uniforme a lo largo de la envergadura. En esta condición se transfiere el mínimo de energía cinética a la capa de torbellinos para una sustentación dada. Max Munk halló la condición de pérdida mínima.

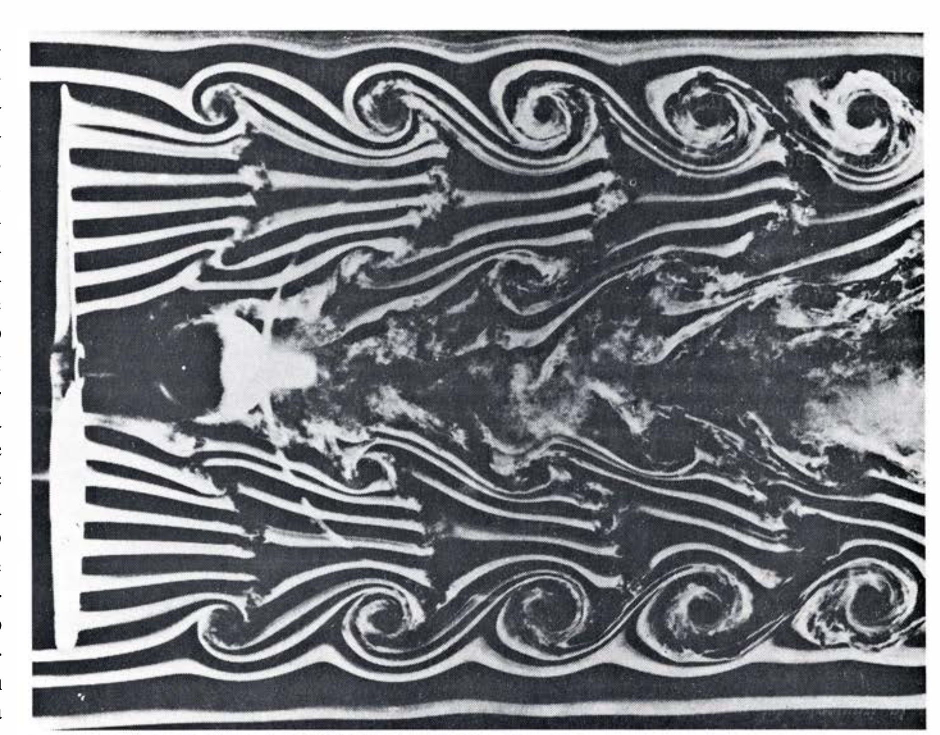
VELOCIDAD DE ROTACION
(Ω)
VELOCIDAD EN EL ESPACIO
(V)
SUSTENTACION POR UNIDAD DE RADIO
$\left(\frac{dL}{dr}\right)$
VELOCIDAD INDUCIDA
(w)
VELOCIDAD DE LA CORRIENTE
(W)
VELOCIDAD AXIAL EN EL ESPACIO
(V)
TORBELLINO DE PUNTA
TORBELLINO DE BUJE
TORBELLINO DE PUNTA
VELOCIDAD DE ROTACION DE LA HELICE EN UN RADIO DADO
(Ωr)
TORBELLINOS LIBRES
DISTRIBUCION RADIAL DE SUSTENTACION
TORBELLINO DE PUNTA
SUSTENTACION POR UNIDAD DE RADIO
$= \frac{dL}{dr} = \rho W\Gamma$
r
TORBELLINO DE BUJE

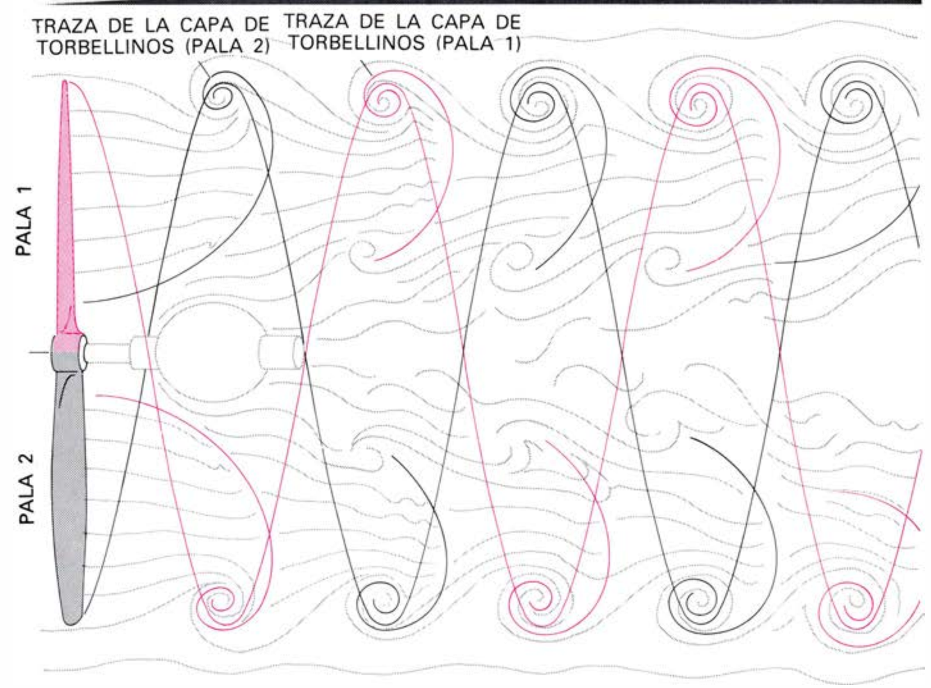
TEORIA TURBILLONARIA DE LAS HELICES. Equivale a la teoría turbillonaria de alas, excepto en que las capas de torbellinos libres de las hélices son superficies helicoidales (*arriba*). Las capas se arrollan en tantos torbellinos de punta cuantas palas tiene la hélice y en un torbellino central de buje, de intensidad igual a la suma de los torbellinos de punta. Los vectores de velocidad inducida (*w*) son perpendiculares a la superficie de las capas helicoidales y corresponden en magnitud al deslizamiento de la teoría de Rankine-Froude sobre la cantidad de movimiento de las hélices náuticas. Para pérdidas inducidas mínimas, cada filamento turbillonario helicoidal (*abajo*) debe moverse perpendicularmente a sí mismo, de tal forma que el conjunto completo de capas de torbellinos parezca moverse axialmente con velocidad de "desplazamiento" constante. El movimiento de la capa de torbellinos para pérdidas mínimas fue descubierto por Albert Betz, colega de Ludwig Prandtl en el Instituto Kalser Wilhelm de Göttingen, Alemania.

mente a la corriente de un fluido real alrededor de una esfera giratoria con superficie rugosa, tal como una pelota de tenis o de golf, que se desplaza y gira simultáneamente, y en sentido opuesto, alrededor de su centro. Se produce sustentación como consecuencia de la aceleración del aire que fluye sobre la parte superior de la pelota y de la deceleración del que fluye bajo ella. De acuerdo con un teorema bien conocido del matemático suizo del siglo XVIII Daniel Bernouilli, el fluido en la corriente acelerada tiene una presión estática más baja que el fluido en la corriente decelerada, de forma que la pelota sufre un empuje de abajo arriba. (Una pelota de tenis que gire en sentido opuesto mientras avanza, recibirá un empuje correspondiente de arriba abajo.) Aunque la corriente de un fluido ideal no viscoso satisface el teorema de Bernouilli, no puede existir sustentación sin circulación. Como resultado de ello, la explicación que suele darse en una escuela secundaria de la sustentación de un perfil en términos del teorema de Bernouilli es totalmente inadecuada.

Según esa explicación, la sustentación se crea porque el aire que pasa sobre la superficie curvada superior de un perfil típico debe recorrer un espacio mayor y, por tanto, debe desplazarse más rápidamente que el aire que pasa bajo la superficie inferior y más plana del mismo. Por consiguiente, la presión sobre el ala debe ser menor que la presión debajo del ala, dando lugar a sustentación. En realidad, la sustentación puede ser generada por un perfil perfectamente simétrico que esté inclinado hacia arriba en un ángulo de ataque adecuado. ¿Cómo sabe la corriente en qué forma separarse en el borde de ataque y reunirse de nuevo en el de salida, de manera que el movimiento sobre la superficie superior del ala tenga una velocidad media más alta y una presión más baja que el movimiento bajo ella?

El ángulo de ataque es por tanto crucial, pero no hay teoría física satisfactoria para explicar por qué se divide la corriente precisamente en la forma en que lo hace. El ajuste de Joukowsky de la intensidad del torbellino para producir un punto de remanso en el punto singular posterior del cilindro transformado (esto es, el perfil) tiene el efecto de alinear el campo de corriente con el borde de salida del perfil, una suposición plausible llamada hipótesis de Joukowsky o hipótesis de Kutta-Joukowsky. El efecto de alinear la corriente en el borde de salida es el de





ESTELA DE UNA HELICE DE DOS PALAS, una hélice de modelo de avión (*arriba*), fotografiada en un túnel aerodinámico especial a finales de los años 50 por F. M. N. Brown, de la Universidad de Notre Dame. La fotografía muestra cómo los filamentos de humo, de distribución vertical, se curvan en una serie de torbellinos de punta, espaciados uniformemente con el paso de la hélice. El torbellino del buje de la hélice está interrumpido por la carena del dispositivo de accionamiento del árbol. El diagrama inferior muestra que las discontinuidades curvadas corresponden a secciones a través de las capas de torbellinos helicoidales que se desprenden de las palas. Las hélices y alas tienen pérdidas de fricción por viscocidad del medio.

desplazar el punto en que se divide la corriente en el borde de ataque del perfil, de tal forma que aparezca circulación y sustentación. El borde de ataque debe ser bien redondeado, como lo son los perfiles matemáticos de Joukowsky.

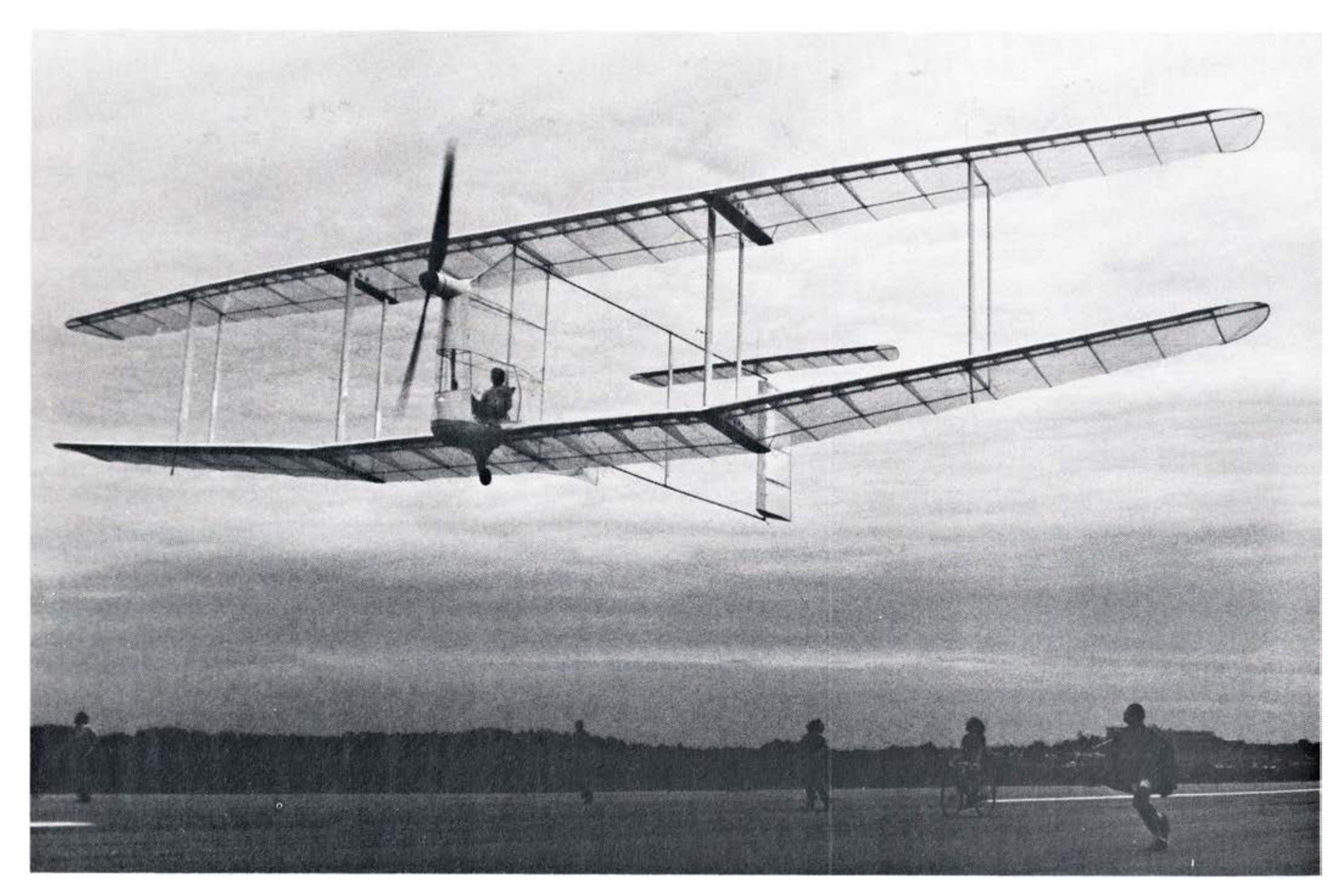
El ángulo de ataque del perfil puede variarse inclinando el eje que une los puntos singulares anterior y posterior de la transformación con respecto a la dirección del campo de velocidades remoto, esto es, su dirección antes de que haya sido influida por el perfil. La curvatura del perfil, o curvatura de su línea central, puede variarse desplazando el radio que une el centro del cilindro con el punto singular posterior con respecto al eje de los puntos singulares en la transformación. Cada una de estas desviaciones da lugar a un cambio en la intensidad de la circula-

ción, o torbellino ligado, y resulta en cambios correspondientes de la sustentación del perfil. Ensayos de los perfiles de Joukowsky en túnel aerodinámico, efectuados por Joukowsky y sus colaboradores en Rusia y por Prandtl y sus colaboradores en Alemania, justo antes de la primera guerra mundial, verificaron las predicciones de la teoría para pequeños ángulos de ataque y pequeños valores de la curvatura. Todas las teorías modernas del perfil incorporan los resultados de Joukowsky como caso particular.

Aunque la teoría de Joukowsky explica la sustentación de secciones del ala en movimiento bidimensional (esto es, en ausencia de componentes de la velocidad de la corriente en dirección de la envergadura) presenta dificultades cuando se aplica a la corriente alrededor de alas de envergadura finita, en las cuales existe corriente en la dirección de la envergadura y donde la sustentación y el torbellino ligado deben descender suavemente hasta cero en los bordes marginales del ala. Este problema fue resuelto por Prandtl, Betz, Max Munk y C. Wieselberger, en Göttingen, poco antes o durante la primera guerra mundial. Introdujeron el concepto de la capa de torbellinos libres compuesta de líneas infinitesimales de torbellinos aproximadamente alineados con la dirección de vuelo. La intensidad de cada uno de los torbellinos libres es proporcional al gradiente local en dirección de la envergadura del torbellino ligado, asociado con el valor local de sustentación del perfil en el modelo de Joukowsky.

Los torbellinos libres son especialmente intensos en las proximidades de los bordes de la capa, en los bordes marginales del ala, donde el gradiente tiende hacia infinito. La capa de torbellinos libres se arrolla hacia arriba alrededor de sus intensos bordes exteriores, formando dos "torbellinos de punta" distintos (que, sin embargo, no fueron tenidos en cuenta en la teoría primitiva). El proceso de arrollamiento se completa sustancialmente varias envergaduras por detrás del borde de salida, y los torbellinos de punta quedan separados menos de una envergadura. Cuando la humedad de la atmósfera está dentro de un cierto margen, el vapor de agua del escape de los motores del avión se condensa dentro de los núcleos de los torbellinos de punta y los hace visibles, como estelas de condensación.

El campo colectivo de velocidades de los torbellinos sobre la capa de los mismos da lugar a una velocidad inducida que tiende a deflectar la capa hacia abajo en cada fase del proceso de arrollamiento. En particular, si la sustentación del ala está distribuida de forma elíptica a lo largo de la envergadura, la capa de torbellinos (antes de su arrollamiento) se deflecta hacia abajo con velocidad uniforme a lo largo de toda su anchura. La energía cinética asociada con el campo de velocidades debido a los torbellinos libres es irrecuperable, pero las pérdidas resultan mínimas para una sustentación, velocidad de vuelo y envergadura del ala dadas, si la distribución de sustentación es elíptica. Este importante resultado se atribuye a Munk.



BIPLANO A PEDAL "CHRYSALIS", construido por estudiantes del Instituto de Tecnología de Massachusetts en junio del pasado año. Se escogió la configuración en biplano para minimizar la resistencia. La envergadura de las alas hubo de limitarse a 22 metros por el espacio del hangar. Dada su pequeña envergadura y su peso de 43 kilogramos (un 30 por ciento más que el *Gossamer Albatross*) *Chrysalis* requería una aplicación de 0,3 a 0,4 CV a los pedales, por lo menos y según el peso del piloto. La hélice del *Chrysalis*, como la del *Gossamer Albatross*, se diseñó con un programa de ordenador preparado por Hyong Bang (uno de los estudiantes, junto con Parks y Youngren), a partir de un algoritmo ideado por el autor. La fotografía es de S. Finberg.

El modelo de Joukowsky para la sustentación de un perfil en corriente bidimensional puede incorporarse a una teoría turbillonaria para alas de envergadura finita calculando la magnitud de la velocidad inducida por los torbellinos libres del sistema. La suma vectorial de la velocidad inducida y la velocidad de vuelo crea un ángulo de ataque inducido. El ángulo de ataque geométrico necesario con respecto a la corriente no perturbada se obtiene sumando el ángulo de ataque inducido al ángulo de ataque teórico necesario para un ala de envergadura infinita en movimiento bidimensional (Nicolai Joukowsky).

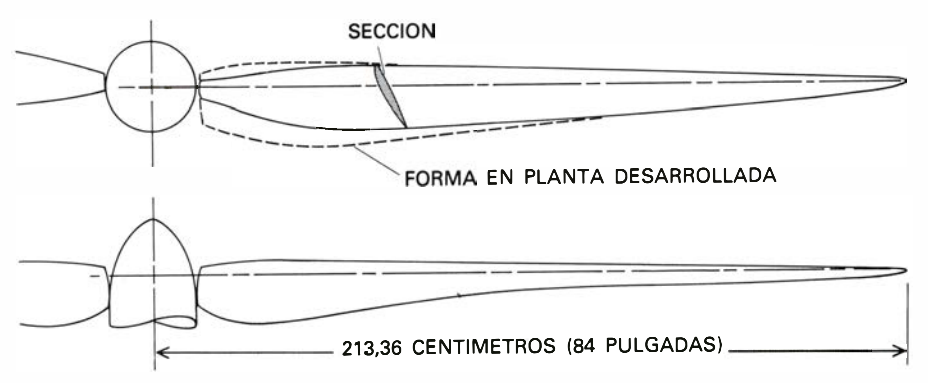
Los vectores de sustentación del ala están también girados hacia atrás un ángulo igual al ataque inducido, creando con ello una resistencia inducida. Por último, la pérdida de energía por unidad de tiempo que corresponde al producto de la resistencia inducida por la velocidad de vuelo equilibra exactamente a la energía cinética añadida continuamente al sistema de torbellinos. Existe además una resistencia de fricción sobre las secciones del ala debida al efecto de viscosidad en las capas límite adyacentes a la superficie del ala, un efecto que fue también descrito por Prandtl. Las predicciones de la teoría turbillonaria del ala fueron confirmadas mediante ensayos en túnel aerodinámico en Göttingen durante la primera guerra mundial.

A partir de aquí solamente existe un paso hasta la teoría turbillonaria de la hélice. La diferencia fundamental consiste en que las capas de torbellinos de la hélice son inicialmente helicoidales. Después se arrollan en tantos torbellinos de punta como palas tiene la hélice y en un torbellino central de buje de rotación opuesta y cuya intensidad es igual a la suma de las intensidades de los torbellinos de punta individuales. El campo colectivo de velocidades de las capas helicoidales de torbellinos es idéntico a la estela de la hélice en funcionamiento. Y, así como existe un único movimiento de la capa de torbellinos que minimiza las pérdidas inducidas del ala sustentadora, existe también un único movimiento debido a la capa de torbellinos que minimiza las pérdidas inducidas de una hélice generadora de empuje. Este movimiento fue descubierto por Betz, y la distribución radial correspondiente (pero aproximada) de vorticidad ligada para un modelo de Joukowsky de distribución óptima de sustentación sobre la pala fue calculada por Prandtl. Publicaron sus resultados en 1919. Goldstein se unió al grupo de Göttingen durante los años 20 y en su tesis doctoral, fechada en 1929, definió los errores en la distribución de circulación aproximada de Prandtl.

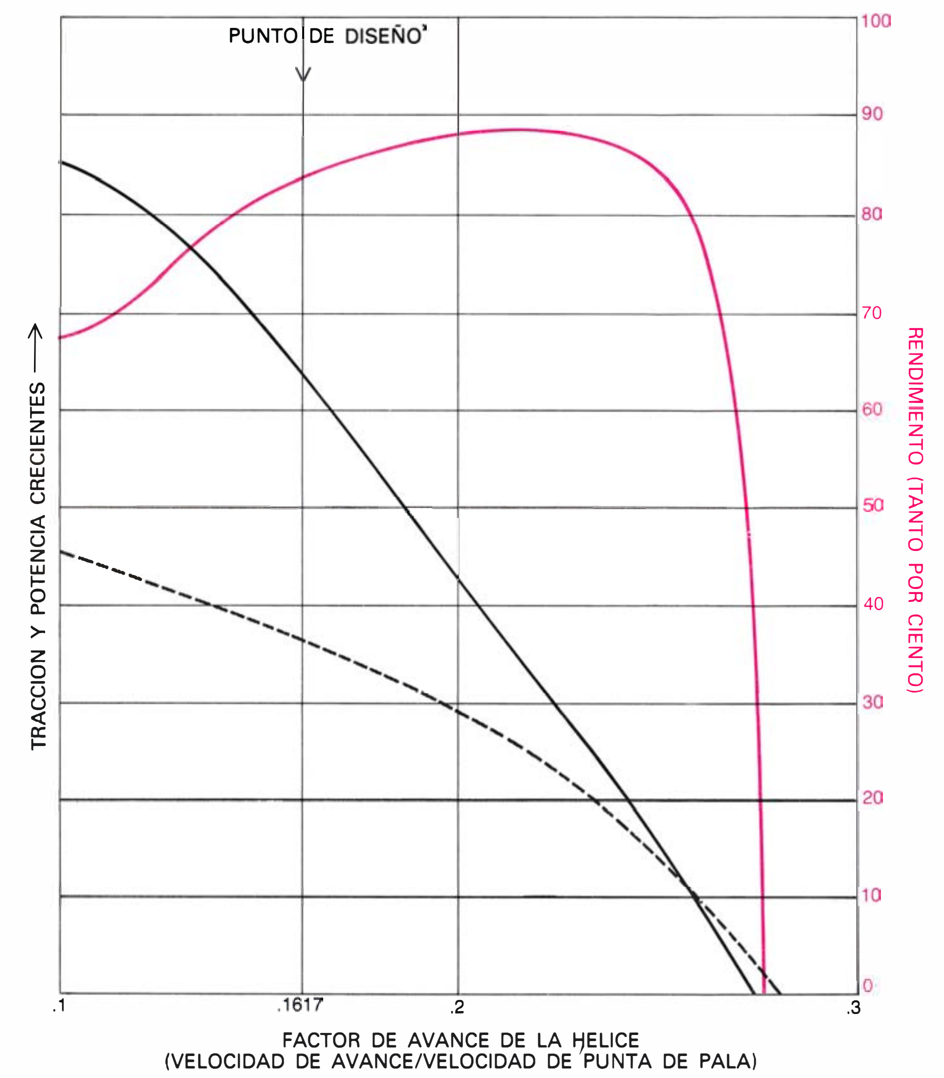
Quizá sea justo decir que incluso ingenieros aeronáuticos han tenido dificultades en asimilar las contribuciones de Betz, Prandtl y Goldstein, justamente a causa de que las capas de torbellinos de una hélice y su desarrollo geométrico en el espacio son difíciles de visualizar en tres dimensiones. La hélice es, después de todo, un tipo peculiar de ala. A diferencia del ala, que en vuelo rectilíneo encuentra el aire en todas partes a la misma veloci-



"GOSSAMER ALBATROSS" aproximándose a la costa francesa en junio del año pasado. Alcanzó así el premio de 100.000 libras ofrecido por Henry Kremer, un industrial británico. Allen, el piloto, de 62 kilogramos de peso, es un ciclista profesional de 27 años y un entusiasta del vuelo en delta. La travesía, de 35,82 kilómetros, se realizó con viento en contra en un tiempo de 2 horas y 49 minutos. El aparato no superó los 4,5 metros de altura sobre el agua. El instrumental de Allen estaba compuesto por un anemómetro y un altímetro dotado de un enfocador sonárico. La velocidad media del pedaleo fue de 75 revoluciones por minuto, equivalentes a un rotación de la hélice de 95 revoluciones por minuto. La fotografía fue obtenida por Don Monroe.



HELICE DEL "CHRYSALIS" en planta (*arriba*) y en alzado (*abajo*). La "forma en planta desarrollada", indicada por la línea de puntos, es la geometría que tendría la pala si no tuviese torsión. La cuerda máxima está a 30 por ciento del radio. El diámetro de la hélice, de 4,27 metros es alrededor del 4 por ciento mayor que la del *Gossamer Albatross*. Ambas hélices se diseñaron de acuerdo con los algoritmos del autor para pérdidas inducidas mínimas. *Chrysalis* se proyectó para una potencia de 0,5 CV (373 watt) en el árbol, una velocidad de rotación de 135 revoluciones por minuto, una velocidad de vuelo de 4,88 metros por segundo (10,9 millas por hora) y, por tanto, un "factor de avance" (velocidad de avance dividida por velocidad de punta de pala) de 0,1617. Debido a su elevada velocidad de punta de pala, tales hélices resultan inadecuadas para aviones que vuelen a más de 50 metros por segundo (112 millas por hora).



COMPORTAMIENTO TEORICO de la hélice de *Chrysalis*. Los coeficientes de empuje (*Línea negra de trazo continuo*), potencia (*línea negra a trazos discretos*) y rendimiento total (*color*) varían con el factor de avance de la hélice. Las pérdidas inducidas mínimas para la carga discal del diseño tienen lugar en el punto de diseño, en el que el rendimiento de la hélice se aproxima al 83 por ciento. El rendimiento continúa aumentando a factores de avance mayores (equivalentes a girar la hélice más despacio a una velocidad de vuelo dada), aunque las pérdidas inducidas ya no resulten minimizadas; se hacen menores conforme se reduce la carga de la hélice y la velocidad axial media de la estela disminuye. Los coeficientes de tracción y de potencia son mayores cuando la hélice gira más rápida que en el punto de diseño (*curvas a la izquierda del punto de diseño*) y disminuyen a medida que la hélice gira más despacio.

dad media, y tiene así un ángulo de ataque común desde el encastre hasta la punta, la hélice topa con el aire a velocidad distinta en cada radio. Es por esto por lo que la pala de hélice se diseña con torsión. Cerca del encastre, donde la velocidad axial, o hacia delante, de la sección de pala es grande comparada con la velocidad circunferencial, la cuerda de la pala se separa sólo ligeramente en la dirección de vuelo. En la punta de la pala, donde la velocidad circunferencial es elevada comparada con la velocidad hacia adelante, la cuerda puede formar un ángulo de torsión casi perpendicular a la línea de vuelo.

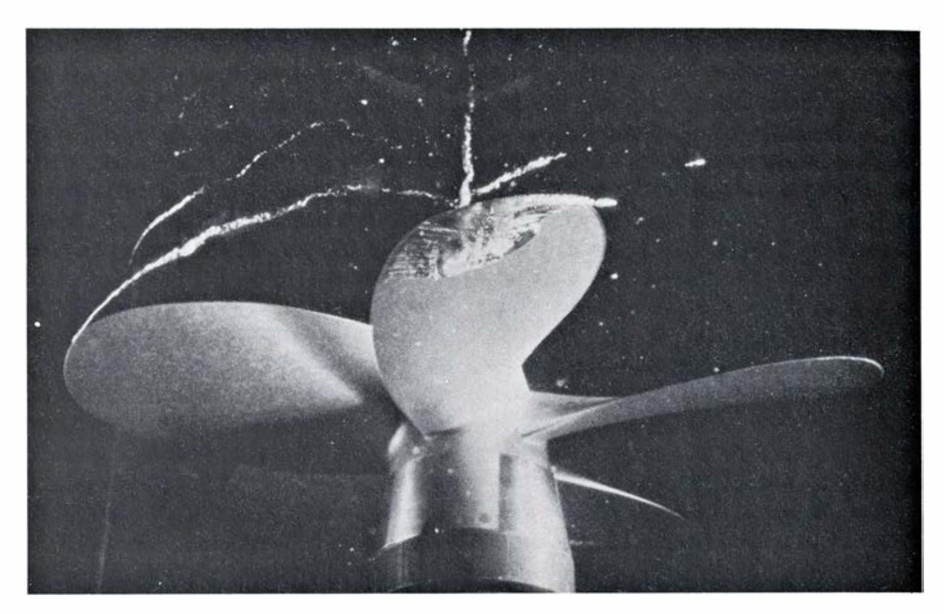
Como hemos visto, el campo de velocidades debido a la capa de torbellinos libres crea una velocidad inducida que es perpendicular a la dirección del movimiento del perfil en el espacio no perturbado. Cerca del buje de la hélice, por tanto, la velocidad inducida está inclinada sólo ligeramente hacia atrás con respecto a la línea de vuelo. En otras palabras, el vector velocidad inducida tiene cerca del buje una componente rotacional elevada y una componente axial pequeña. De acuerdo con ello, las secciones de pala cerca del buje contribuyen sólo en pequeña parte en la tracción de la hélice hacia adelante. Cerca de la punta de la pala, por el contrario, la velocidad inducida tiene una componente axial elevada y una componente rotacional mucho más pequeña. Como resultado, el empuje por unidad de radio para una pala de sección constante aumentaría desde el buje hasta la punta, excepto cerca de la misma, donde debe tender a cero "elípticamente".

Podría pensarse que, como la parte exterior de la pala realiza un trabajo más efectivo que la parte interior, la exterior debiera aumentarse en el diseño de la hélice. Uno podría imaginar, por ejemplo, una hélice cuyas palas aumentasen de anchura desde el buje hasta la punta. Tal geometría (aun cuando pudiese conseguirse estructuralmente) sería relativamente ineficiente. La contrapartida en la hélice de la distribución uniforme de velocidad inducida para un ala eficiente se consigue cuando la velocidad inducida en cualquier radio de la pala de la hélice es igual a la mitad de la velocidad de la capa de torbellinos en el mismo radio, y cuando las velocidades de la capa de torbellinos, axial y rotacional, varían con el radio de tal forma que presentan el aspecto de una velocidad uniforme axial de desplazamiento. Un anuncio

cilíndrico giratorio de barbero, con franjas helicoidales, por ejemplo, tiene una velocidad de desplazamiento axial finita, pero ninguna velocidad axial física real.

En general, la geometría de la hélice de máxima eficiencia, conocida por geometría de mínimas pérdidas inducidas, es la de una hélice cuyas palas tienen una cuerda máxima alrededor del 30 o 40 por ciento de su radio y que se estrechan después hacia una punta afilada. Tales palas eran una característica común de las hélices de aleación de aluminio fabricadas a finales de los 20 y comienzos de los 30. Conforme los motores se hicieron más potentes y se sobrecargaron para alcanzar mayores altitudes, se transformó en una práctica común el incrementar el área de la pala hacia la punta con objeto de absorber mayor potencia sin aumentar el radio de la hélice. No se utilizaron diámetros mayores, porque las velocidades de punta de pala ya eran transónicas, esto es, se movían aproximadamente a la velocidad del sonido, un régimen para el cual no se conocían diseños satisfactorios de pala. En casos extremos las puntas de pala se hicieron incluso cuadradas. Tales desviaciones de la geometría de mínimas pérdidas inducidas conducían a pequeñas pérdidas de rendimiento de la hélice, que no fueron observadas por la mayoría de los ingenieros aeronáuticos de la época, la mayor parte de los cuales probablemente no habían leído o quizá no habían comprendido el trabajo de Goldstein (o los comentarios al mismo de Hermann Glauert y más tarde de Theodore Theodorsen). Sólo cuando el proyectista no puede prescindir de ninguna contribución al rendimiento, por pequeña que sea, resulta esencial un diseño de mínimas pérdidas inducidas.

Mi propia contribución a la teoría de la hélice ha consistido en desarrollar una teoría "radialmente graduada" de la cantidad de movimiento, que se presta a ser implementada en ordenador y que es consistente con la forma de Betz-Prandtl de la teoría turbillonaria de la hélice. Todas las teorías de la hélice tratan de calcular la velocidad inducida de los elementos de pala que se debe al sistema de torbellinos libres. En el caso de hélices con pérdidas inducidas mínimas, la mitad de la velocidad de desplazamiento corresponde a la velocidad de estela de la teoría del disco actuador de Rankine y Froude. Como en la teoría del ala, las componentes axiales y rotacionales de la velocidad inducida llevan consigo un aumento del ángulo de pala para un valor dado de empuje, junto con una rotación hacia atrás de los vectores de sustentación de pala, lo que incrementa el par necesario para hacer girar la hélice. El incremento requerido corresponde a un momento inducido. La teoría turbillonaria de la hélice es, por tanto, totalmente consistente con las teorías iniciales de cantidad de movimiento para las hélices náuticas y tiene en consideración detalles de distribución de carga sobre la pala que corresponden a una geometría específica de la hélice.



LAS HELICES MARINAS EN ACCION se estudian en el laboratorio de hidrodinámica marina del Instituto de Tecnología de Massachusetts con la ayuda de la fotografía de alta velocidad. En esta fotografía, en que se aprecia la hélice desde arriba, la región blanca que aparece en la punta de la pala central es una gran burbuja de vapor de agua, prueba de una brusca cavitación producida por un campo de corriente no uniforme que simula la estela de la carena de un barco. Las bajas presiones creadas en la superficie de "succión", o anterior, de las palas inducen el hervor instantáneo del agua circundante. Al girar la pala, la gran burbuja se deshace y genera un rastro de burbujas menores. La cavitación cíclica reduce el rendimiento de la hélice, provoca vibraciones y genera una "firma" acústica característica. La cavitación se reduce al incrementar la superficie de las palas y optimizar su geometría.

Las hélices, al igual que las alas, tienen también pérdidas de fricción, debidas a la acción de la viscosidad del aire que fluye cerca de la superficie del perfil: la capa límite. Estas pérdidas se minimizan haciendo que los elementos de pala operen a ángulos de ataque para los cuales la relación resistencia/sustentación sea baja y, si es posible, eligiendo ángulos de hélices iguales a 45 grados menos la mitad del ángulo cuya tangente es la relación de sustentación a resistencia de la pala en dos dimensiones. La reducción de las pérdidas de fricción exige, por tanto, valores bastante elevados del "factor de avance" (relación de la velocidad hacia adelante a la velocidad de punta de pala de la hélice de alrededor de 0,7) y un pequeño número de palas, de forma que la solidez (o relación superficie de pala a superficie de disco) pueda concentrarse en unas pocas palas anchas que operen a números de Reynolds elevados y fricción superficial correspondientemente baja. (El número de Reynolds es la relación de las fuerzas inerciales a las fuerzas de viscosidad en un fluido; ayuda a determinar el carácter de la capa límite y su resistencia a la separación.) Las pérdidas inducidas, por otra parte, se minimizan aumentando el número de palas y disminuyendo el factor de avance con objeto de acortar la separación entre las capas individuales de torbellinos y hacerlas más perpendiculares al eje de la hélice. El objetivo es hacer el campo de velocidades en la estela tan uniforme como sea posible. Las exigencias de bajas pérdidas inducidas y bajas pérdidas de fricción en una hélice son, por tanto, contrapuestas; las hélices de máximo rendimiento exigen un equilibrio óptimo entre las dos.

Utilizando la distribución analítica aproximada de circulación de Betz-Prandtl he conseguido preparar algoritmos simples, o procedimientos de cálculo, que permiten a un proyectista de hélice determinar la geometría de la hélice de mínimas pérdidas inducidas compatibles con cualquier ligadura de diseño que pueda impedir que se alcancen las pérdidas inducidas mínimas

posibles. (Por ejemplo, si el diseño necesita una hélice bipala, por la razón que sea, el proyectista no puede explorar las posibles ventajas de añadir más palas.) Estos algoritmos conducen así a una hélice del máximo rendimiento posible que desarrolle un empuje o absorba una potencia dada para diámetro, número de palas, velocidad de vuelo, velocidad del árbol y densidad del aire dadas. Los algoritmos tienen también en consideración una amplia gama de relaciones de sustentación a resistencia para elementos de pala de diferentes formas, con objeto de conseguir coeficientes de sustentación deseados en puntos de diseño especificados. Como los cálculos dan la eficiencia de la hélice no sólo en el punto de diseño, sino también en condiciones de vuelo apartadas del mismo, pueden ser utilizados por el proyectista para cambiar su punto de diseño.

He tenido la buena fortuna de tomar parte en una demostración práctica de la validez de lo que podría propiamente ser llamada una teoría clásica de la aerodinámica de la hélice, publicada por primera vez en 1919 (un año antes de que yo naciese) y prácticamente olvidada hacia 1942, cuando comencé mi carrera profesional en la planta de la Curtiss-Wright Corporation en Buffalo (Nueva York). Debo esta demostración a unos entusiastas estudiantes del Instituto de Tecnología de Massachusetts (MIT) que decidieron construir un avión accionado a pedales en enero de 1979 y que acudieron a mí en busca de las ideas sobre diseño de hélice que en principio había organizado para presentar en un simposio.

Los estudiantes fueron estimulados por el sueño (remoto, para ser exactos) de ganar el premio de 100.000 libras ofrecido por Henry Kremer, un industrial británico, para un avión accionado por la potencia muscular de una persona, capaz de cruzar el Canal de la Mancha. (Un premio Kremer anterior, de 50.000 libras, para el primer avión propulsado por potencia muscular capaz de completar una trayectoria en forma de ocho alrededor de dos mástiles separados por una distancia de media milla, fue ganado en agosto de 1977 por *Gossamer Condor,* un avión ultraligero diseñado por Paul McCready, presidente de AeroVironment Incorporated, de Pasadena, California.)

Como primer paso, mis estudiantes Hyong Bang, Robert Parks y Harold Youngren codificaron una versión de mi algoritmo de la hélice para cálculo en ordenador y lo aplicaron al rediseño de la hélice para un modelo de vuelo libre radiocontrolado a escala un octavo de un biplano al que pensaban llamar *Crysalis.* El modelo, con una envergadura de 2,74 metros, estaba propulsado por el motor de una cámara Polaroid SX-70, que accionaba una hélice a través de una transmisión con una relación de reducción de 11 a 1. El modelo se construyó en febrero, en una semana. Comparado con una hélice diseñada intuitivamente, con la cual el modelo apenas se mantenía en vuelo horizontal, con las baterías totalmente cargadas, la hélice teóricamente correcta, adaptada a la estructura y a la planta de potencia, permitió al modelo volar tres figuras y media en forma de ocho alrededor de mástiles a veinte metros de distancia en el gimnasio del Instituto de Tecnología de Massachusetts, antes de que las baterías se descargasen.

El segundo éxito para la revivida teoría de Betz-Prandtl llegó cuando los mismos estudiantes diseñaron una nueva hélice para el monoplano *Gossamer Albatross,* que el grupo de McCready había construido para su propio intento de obtener el premio de cruce del Canal. Bryan Allen, un joven ciclista profesional, que pesaba 62 kilogramos, había sido capaz de mantener el *Gossamer Albatross* en el aire solamente durante 17 minutos, cuando estaba equipado con su hélice original. Con la nueve hélice MIT, en combinación con mejoras en la estructura del avión, Allen estimó que podría mantener el avión en el aire durante horas, o hasta que su propio "suministro de combustible" de glucosa y agua se agotase. El cruce real del Canal, obstaculizado por vientos de frente, exigió que Allen pedalease durante 2 horas y 49 minutos. La distancia recorrida fue de 22,26 millas; la velocidad media con

"PROPFAN" TRANSONICA. Se trata de uno de los varios proyectos que tienen en cartera la United Technologies Corporation y la National Aeronautics and Space Administration para un posible servicio en aviones de turbopropulsión (turboprop) capaces de alcanzar las velocidades de los aviones turbofan. Si el ingenio tiene éxito consumirá un 20 por ciento menos de combustible que los modelos actuales. El propfan se denomina transónico porque la velocidad de las puntas de las hélices excederá la velocidad del sonido en un 13 por ciento. El avión volará a un 80 por ciento de la velocidad del sonido (852 kilómetros por hora a 10.600 metros). Para operar a régimen transónico una hélice de avión debe acercarse a la elevada solidez de las hélices náuticas. La solidez de esta hélice es del 70 por ciento. Las palas anchas proporcionan las elevadas relaciones de sustentación/resistencia necesarias para operar a niveles transónicos. Se ha elegido la forma curvada de las palas para reducir el "ruido de campo próximo" (en el interior de la cabina) que viene causado por la centralización del campo de presión en rotación.

respecto al suelo fue de 7,9 millas por hora.

El tercer éxito para la teoría de Betz-Prandtl llegó con el diseño de la hélice para el *Crysalis* de tamaño real. Construido entre marzo y junio, el biplano efectuó más de 320 vuelos durante el verano antes de que fuese finalmente desmantelado en septiembre. Si *Crysalis* no hubiese tenido una hélice teóricamente correcta, acoplada a la estructura y planta de potencia (0.5 CV), dudo que con su pequeña envergadura y substancial peso (10 kilogramos más que *Gossamer Albatross*) hubiese podido volar en absoluto.

Comparados con los problemas del proyectista de hélices de avión, los problemas del proyectista de hélices de barco tienen muchas más ligaduras. El diámetro de la hélice está limitado por una parte por el máximo calado permisible del buque cargado y por otra parte por su línea de flotación cuando está poco cargado. La hélice debe operar en un campo de corriente fuertemente no uniforme, distorsionado por el contorno de la popa y muchos defectos de velocidad producidos por la capa límite del casco. A pesar de una geometría multipala favorable, la fuerte limitación en el diámetro de la hélice conduce a una carga discal excesiva y a pérdidas inducidas importantes. La solidez de la hélice, que ya de por sí es grande debido a la elevada carga discal, se incrementa aún más a causa de los límites impuestos en el ángulo de ataque de la pala y en su curvatura por la necesidad de limitar la cavitación en la corriente. La cavitación se produce cuando la velocidad del fluido cerca de las palas se eleva tanto que la presión local disminuye por debajo de la presión de vapor del agua. En este caso el agua hierve, creando repentinamente cavidades en el campo de corriente. La cavitación es la principal fuente de ruidos de las hélices marinas.

Se presentan difíciles problemas adicionales debido a frecuencias resonantes de excitación que pueden establecerse entre la pala giratoria y el casco del buque. Generalmente, las palas de hélices náuticas se construyen con bordes de ataque con flecha positiva a fin de hacer que la excitación sea menos aguda como función del ángulo de rotación de pala. Por todas estas razones, los análisis de hélices de buque deben tener en cuenta la variación de velocidad inducida sobre la superficie total de la pala, así como el arrollamiento de la capa de torbellinos, que tiene lugar muy rápidamente para hélices muy cargadas. Las distribuciones aproximadas simples de velocidad inducida y de circulación de Betz Prandtl y Goldstein no son aplicables a las hélices de buques; el proyectista debe recurrir a métodos numéricos semiempíricos de análisis. Al final, la eficiencia de las hélices náuticas es usualmente de alrededor del 70 por ciento. Los ingenieros navales aceptan esto como lo mejor que puede conseguirse, dada la naturaleza de las fuertes ligaduras prácticas.

En la actualidad se está realizando un extenso estudio aplicado a hélices para aviones que se pretende que vuelen al 80 por ciento de la velocidad del sonido, lo que corresponde a unos 850 kilómetros por hora, a altitudes de crucero de 10,7 a 13,7 kilómetros. Las palas de tales hélices tendrán también velocidades rotacionales de punta de pala del 80 por ciento de la velocidad del sonido, y cuando se añada la componente de velocidad hacia adelante, la velocidad de punta de pala alcanzará el 113 por ciento de la velocidad del sonido. A tales velocidades, las hélices habrán de ser multipalas, a fin de conseguir buenas relaciones de sustentación a resistencia, y tendrán solidez elevada. En consecuencia, se parecerán mucho a las hélices marinas. Se deberán alcanzar rendimientos del 80 por ciento. Las hélices de alta velocidad serán ruidosas en operación (por razones muy diferentes de las que provocan el ruido de las hélices náuticas). El ruido, sin embargo, será principalmente ruido de "campo próximo", que podría rebajarse a niveles tolerables mejorando el aislamiento acústico del fuselaje. El ruido de "campo lejano" puede ser realmente más pequeño que el de los motores turbofan existentes. La eficiencia en cuanto a combustible de un motor turbohélice, dada una hélice de alta velocidad adecuada, debiera ser alrededor del 20 por ciento mayor que la eficiencia de un mctor turbofan de alto desvío ("high-by-pass") con un "generador de gas" (la combinación de compresor, cámara de combustión y turbina de potencia) del mismo tamaño que el que sería necesario para el motor turbohélice. La hélice del motor turbo-hélice tendría un diámetro aproximadamente doble que el ventilador del motor turbofan.

Tras más de un siglo de aplicaciones en hidronáutica y aeronáutica, la hélice sigue viva y girando; supongo que perdurará en tanto el hombre utilice el agua y el aire para el transporte.

Juegos matemáticos

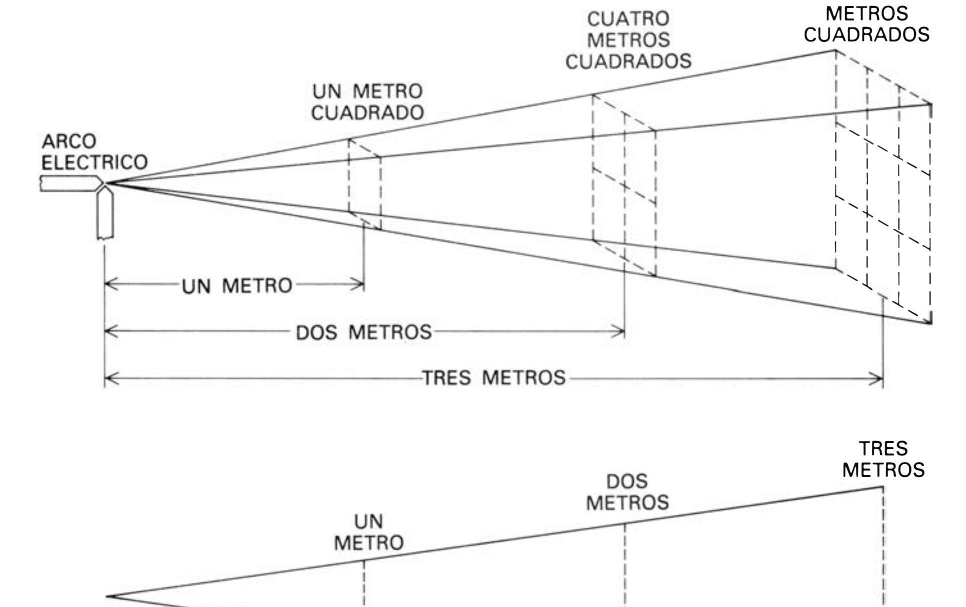
Los placeres de crear ciencia y técnica en el planiverso ideal de los geómetras

Martin Gardner

"Los científicos que se ocupan de planiversos son ralea poco común."
ALEXANDER KEEWATIN DEWDNEY

Que sepamos, el único universo que existe realmente es el que nos acoge, provisto de tres dimensiones espaciales, y de otra dimensión más, el tiempo. No cuesta mucho imaginar, y numerosos autores de ciencia ficción así lo han hecho, que en espacios de cuatro dimensiones pudieran existir organismos inteligentes; en cambio, con dos dimensiones hay tan pocos grados de libertad que desde hace mucho se ha presumido imposible la existencia de formas bidimensionales de vida inteligente. Se han realizado, no obstante, dos notorios intentos dedicados a describir organismos de esta categoría.

Un clérigo londinense, Edwin Abbott Abbott, publicó en 1884 una novela de intención satírica titulada *Flatland* ("Planilandia"; existe traducción castellana en Ediciones Guadarrama; Madrid, 1976). Aunque desdichadamente esta "planilandia" nos deja casi por completo a oscuras sobre las leyes físicas que gobiernan la vida de sus habitantes, tal situación mejoró mucho en 1907, con la publicación de *An Episode of Flatland,* de Charles Howard Hinton. Los personajes y el estilo literario de la narración seguirían faltos de relieve aunque su mundo fuese tridimensional, pero el cuento de Hinton aportó los primeros indicios de la ciencia y técnica posibles para seres bidimensionales. Tan extravagante obra hace mucho que está agotada, y es lástima; de todas maneras, el lector podrá documentarse un poco en el capítulo "Flatlands" de mi libro *The Unexpected Hanging and Other Mathematical Diversions* (Simon and Schuster, 1969).



Ley cuadrática inversa, válida en nuestro mundo (**arriba**) *y su homóloga lineal* (**abajo**) *del mundo plano*

En "Flatlands" escribía yo: "Sería divertido especular cuál sería la física posible, y cuáles los tipos de mecanismos sencillos realizables en un mundo bidimensional". Esta observación atrajo el interés de Alexander Keewatin Dewdney, un especialista en ciencias de cómputo de la Universidad de Western Ontario. Algunas de sus especulaciones más tempranas sobre este tema fueron dadas a conocer en 1978, en un informe interno de la Universidad, y más tarde, en 1979, en un artículo titulado "Exploring the Planiverse", en el *Journal of Recreational Mathematics* (vol. 12, n.º 1, págs. 16-20; septiembre). También en 1979, aunque más tarde, Dewdney publicó por cuenta propia su *tour de force:* 97 páginas dedicadas a ciencia y técnica en un mundo plano, tituladas "Two-Dimensional Science and Technology". Aunque cueste creerlo, lo que verdaderamente hace Dewdney en su libro es sentar los cimientos de lo que él denomina "planiverso": un mundo bidimensional cuya existencia no sea imposible desde el punto de vista lógico. Provisto de leyes físicas, químicas, astronómicas y biológicas propias, el planiverso es análogo cercano de nuestro propio universo (que Dewdney llama "estiriverso") y, según parece, no contiene contradicciones. Es preciso añadir que semejante proeza no pasa de ser el entretenimiento de un matemático cuyos trabajos, digamos, serios, han aparecido en más de 30 artículos de revistas especializadas.

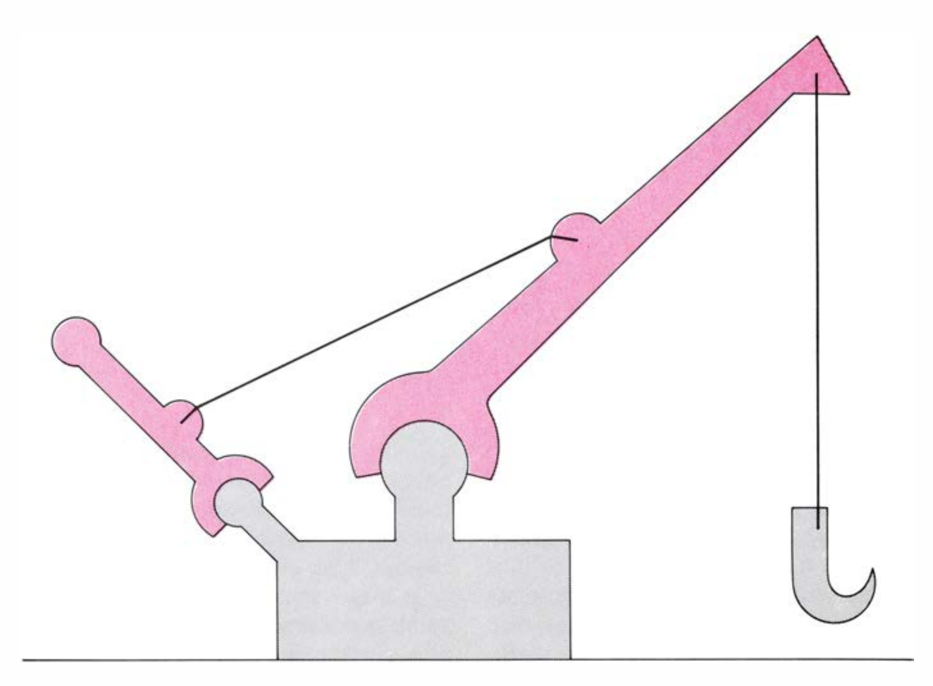
El planiverso de Dewdney recuerda al de Hinton en que tiene una tierra que Dewdney llama, como hiciera ya Hinton, Astria. Astria es un planeta discoidal que gira sobre sí mismo en un espacio planar. Los astrianos, que ca-

minan erguidos por el borde de su planeta, saben distinguir este y oeste, arriba y abajo, pero naturalmente no hay norte ni sur. El "eje" de Astria es un punto situado en el centro del planeta circular. Podemos concebir al planeta plano como verdaderamente bidimensional, o si se quiere, podemos darle un espesor muy fino, e imaginar que se mueve sin fricción entre dos planos paralelos.

Lo mismo que en nuestro mundo, en un planiverso la gravitación es una fuerza de atracción entre los cuerpos, que varía de forma directamente proporcional al producto de sus masas, pero que es inversamente proporcional a la distancia que los separa, en lugar de serlo a su cuadrado. Admitiendo que en el planiverso las fuerzas de tipo gravitatorio o luminoso se muevan en línea recta, es fácil ver que la intensidad de tales fuerzas tiene que ser inversamente proporcional a la distancia. Vemos en la parte superior de la ilustración de la página opuesta la figura que suelen dar los textos para demostrar que la intensidad luminosa varía inversamente con el cuadrado de la distancia; en la parte inferior se muestra la analogía plana correspondiente.

Para que su fantástico proyecto "no degenere en especulación ociosa", Dewdney adopta dos principios fundamentales. El "principio de semejanza" estipula que el planiverso debe ser tan parecido al esteriverso como sea posible. Así, los movimientos no perturbados por fuerzas exteriores deben ser rectilíneos y uniformes, los análogos bidimensionales de las esferas son los círculos, etcétera. El "principio de modificación" enuncia que en aquellos casos donde sea obligado elegir entre hipótesis en conflicto, ambas igualmente homólogas de teorías esteriversales, se deberá elegir la más fundamental, modificando la otra. Para decidir qué hipótesis es más fundamental, Dewdney establece que en su orden jerárquico la física es más fundamental que la química, ésta, más que la biología, y así sucesivamente.

Para aclarar las interconexiones entre los distintos niveles de la teoría, Dewdney examina cómo ha evolucionado la grúa en el planiverso, que vemos esquematizada en la ilustración superior de esta página. El ingeniero proyectista empezó dotándola de brazos más finos que los de la ilustración, pero cuando un metalúrgico le hizo notar que los materiales planares se rompen más fácilmente que sus homólogos tridimensionales, el ingeniero reforzó los brazos. Pero entonces, un químico, invocando los principios de semejanza y modificación a un nivel más profundo, constató que las fuerzas intermoleculares planiversales son mucho más enérgicas de lo que se pensaba, y por ello, el ingeniero volvió a diseñar unos brazos finos.



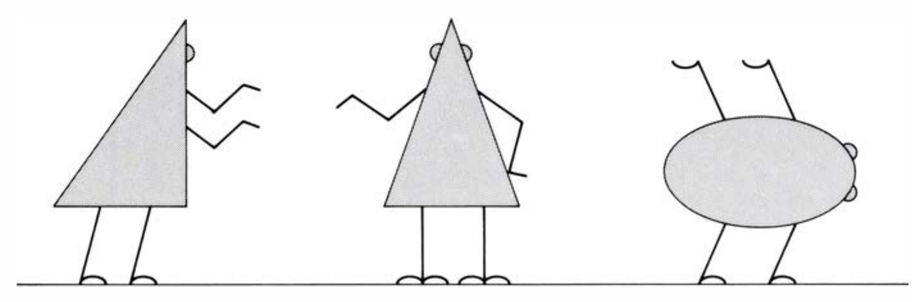
Grúa planiversal

El principio de semejanza lleva a Dewdney a postular que el planiverso es un continuo tridimensional de espacio y tiempo, que contiene materia compuesta por moléculas, átomos y partículas fundamentales. La energía se propaga por ondas, y está cuantificada. Existe luz de todas las longitudes de onda, luz que se refracta por lentes planares. Gracias a ello en el planiverso hay ojos, telescopios y microscopios. El planiverso comparte con el esteriverso principios fundamentales, como los de causalidad, los principios de la termodinámica y las leyes sobre inercia, trabajo, rozamiento, elasticidad y magnetismo.

Dewdney supone que su planiverso nació de una gran explosión (homóloga del "big bang") y que en la actualidad se encuentra en expansión. Un cálculo elemental, basado en la ley de gravitación planiversal (inversamente proporcional a la distancia), muestra que, independientemente de la cantidad de masa del planiverso, la expansión llegará a detenerse, comenzando entonces

NUMERO ATOMICO	NOMBRE	SIMBOLO	ESTRUCTURA ELECTRONICA 1s	2s	2p	3s	3p	3d	4s	4p . . .	VALENCIA
1	HIDROGENO	H	1								1
2	HELIO	He	2								2
3	LITROGENO	Lt	2	1							1
4	BEROXIGENO	Bx	2	2							2
5	FLUORONO	Fl	2	2	1						3
6	NEOCARBONO	Nc	2	2	2						4
7	SODALINIO	Sa	2	2	2	1					1
8	MAGNILICONIO	Mc	2	2	2	2					2
9	ALUSFORO	Ap	2	2	2	2	1				3
10	SULFICONIO	Sp	2	2	2	2	2				4
11	CLOFORO	Cp	2	2	2	2	2	1			5
12	ARGOFURIO	Af	2	2	2	2	2	2			6
13	HINTONIO	Hn	2	2	2	2	2	2	1		1
14	ABBOGENIO	Ab	2	2	2	2	2	2	2		2
15	HAROLDIMIO	Wa	2	2	2	2	2	2	2	1	3
16	LAURANIO	La	2	2	2	2	2	2	2	2	4

Tabla periódica de los 16 primeros elementos planiversales



Astrianos, según Hinton (**izquierda**) *y Dewdney* (**centro**). *Monstruo astriano de ojos saltones* (**derecha**)

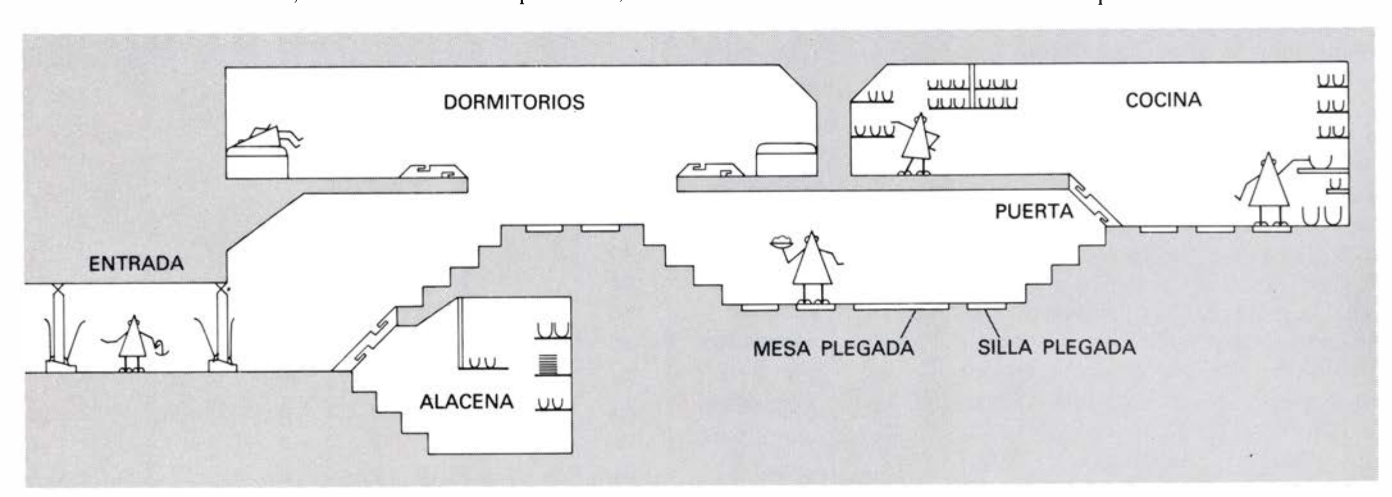
una fase de contracción. El firmamento nocturno astriano será, desde luego, un semicírculo a lo largo del cual están dispersos rutilantes puntos luminosos. Si las estrellas tuvieran movimientos propios, se eclipsarían constantemente unas a otras. Y si Astria tuviera otro planeta hermano, al cabo de cierto tiempo éste habría eclipsado todas las estrellas del firmamento.

Podemos suponer que Astria tiene un movimiento de traslación alrededor de un sol, y que gira sobre sí mismo para crear noche y día. Dewdney descubrió que, en un planiverso, las únicas órbitas estables que retrazan constantemente una misma trayectoria son las circunferencias perfectas. También son posibles otras órbitas estables, sensiblemente elípticas, pero el eje mayor de estas elipses va avanzando, de forma que la órbita nunca se cierra exactamente. Queda por esclarecer si la gravitación planiversal consentiría la existencia de una luna con órbita estable alrededor de Astria. La dificultad se debe a la gravitación solar, y para resolver esta cuestión es necesario trabajar en el análogo planar del problema que nuestros astrónomos conocen por problema de los tres cuerpos.

Dewdney analiza con detalle la naturaleza del clima astriano, con las análogas de nuestras estaciones, vientos, nubes y lluvias. Los ríos astrianos son indistinguibles de los lagos, excepto porque sus corrientes pueden ser más rápidas. Una característica peculiar de la geología astriana es que el agua no puede escurrirse alrededor de las piedras, como en nuestra tierra. En consecuencia, el agua de lluvia se acumula constantemente tras cada una de las piedras de las pendientes, tendiendo a hacerlas rodar ladera abajo. Cuanto más suave sea la pendiente tanta más agua queda retenida y tanto mayor es el empuje. Dewdney concluye que, de haber lluvias periódicas, la superficie astriana sería sorprendentemente lisa y uniforme. Otra de las consecuencias de la imposibilidad de que el agua fluya por los costados es que ésta quedaría atrapada en grandes bolsas en el suelo, tendiendo a crear zonas de traidoras arenas movedizas en los huecos del planeta. Es de esperar, escribe Dewdney, que las lluvias sean poco frecuentes en Astria. Asimismo, los vientos tendrían allí efectos mucho más graves que en la Tierra, pues al igual que las aguas, no pueden pasar en torno a los obstáculos.

Dewdney dedica muchas páginas a construir para su planiverso una química plausible, utilizando como modelo la materia tridimensional y las leyes de la mecánica cuántica. La ilustración inferior de la página anterior muestra la tabla periódica para los 16 primeros elementos planiversales. Dado que los primeros son casi idénticos a los equivalentes de nuestro mundo, se llaman también hidrógeno y helio. Los 10 siguientes tienen nombres compuestos que recuerdan a los elementos esteriversales más semejantes a ellos, por ejemplo, el litrógeno reúne en sí propiedades del nitrógeno y del litio. Los cuatro siguientes reciben sus nombres en honor de Hinton, Abbott y los jóvenes enamorados de la novela de Hinton, Harold Wall y Laura Cartwright.

En el mundo plano los átomos se pueden combinar de modo natural, formando moléculas, pero evidentemente tan sólo serán permisibles enlaces representables mediante grafos planares. (Este resultado traduce la propiedad esteriversal análoga, pues no existen enlaces químicos que se corten.) Lo mismo que en nuestro mundo, puede haber moléculas asimétricas que sean imagen especular una de otra, con lo que no podríamos "darle la vuelta" a ninguna de ellas para hacerla idéntica a la otra. Existen analogías muy llamativas entre la química del planiverso y el comportamiento de capas monoatómicas depositadas sobre caras de cristales de nuestro esteriverso. En nuestro mundo, las moléculas pueden agruparse en 230 grupos cristalográficos distintos, pero en el planiverso solamente pueden formarse 17. Me veo obligado a pasar por alto las especulaciones de Dewdney sobre difusión molecular, leyes magnéticas y eléctricas, analogías de las ecuaciones de Maxwell y sobre otros temas demasiado elevados para resumirlos aquí.

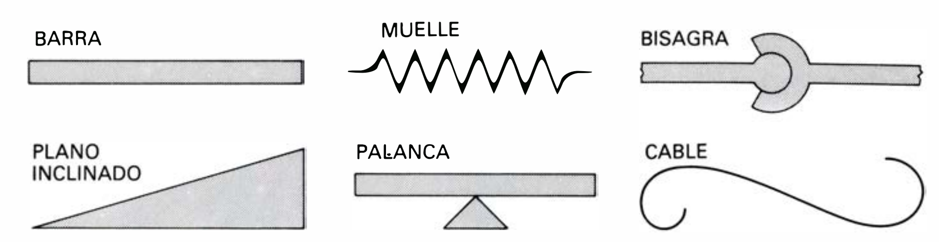


Casa subterránea astriana

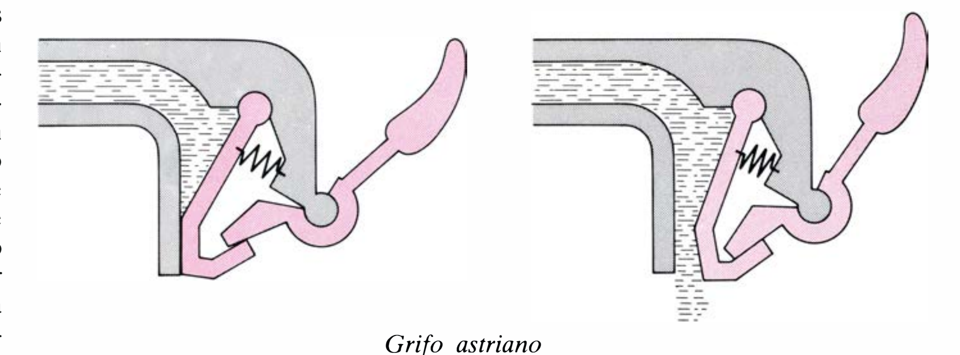
Dewdney supone que los animales de Astria están compuestos por células que se apiñan formando huesos, músculos y tejidos conectivos análogos a los que vemos en biología esteriversal. No encuentra grandes dificultades para mostrar cómo pueden articularse huesos y músculos capaces de crear extremidades móviles que permitan a los animales reptar, caminar, volar o nadar. En realidad, algunos de estos movimientos son más sencillos que en nuestro mundo. Por ejemplo, un animal bípedo esteriversal tiene considerables dificultades para mantenerse en equilibrio mientras camina, en tanto que en el planiverso si un animal tiene ambas patas en tierra no hay forma de hacerlo caer. Además, en el planiverso los seres voladores no pueden tener alas, ni tampoco las necesitan para volar. Si el animal tiene forma aerodinámica, su cuerpo puede actuar de ala (pues el aire solamente puede rodearlo bifurcándose en el plano). Los seres voladores podrían impulsarse mediante colas vibrátiles.

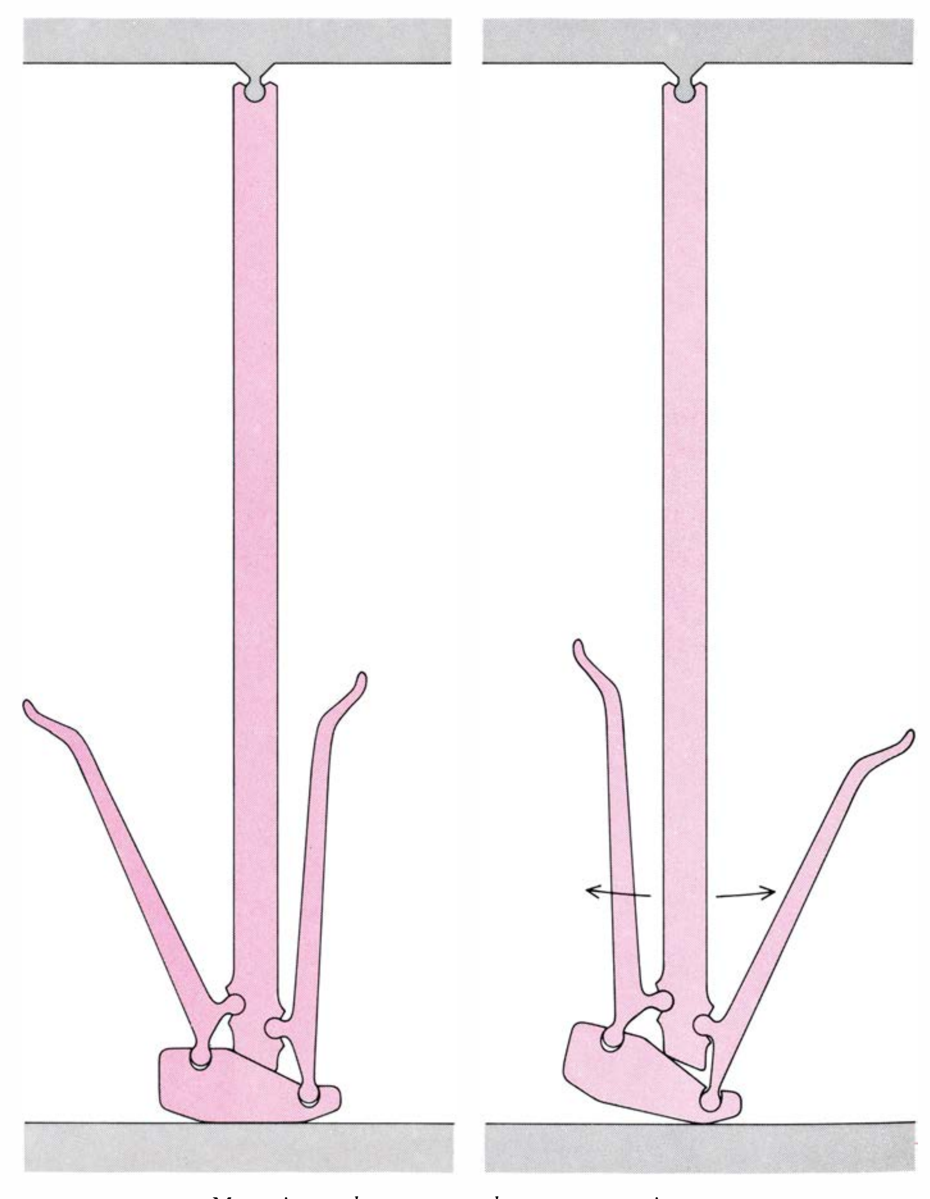
Los cálculos indican que los animales astrianos tendrían también una tasa metabólica probablemente muy inferior a la de los terrestres, porque en comparación poco calor puede escapar a través del contorno de su cuerpos. Además, los huesos de los animales astrianos podrían ser mucho más delgados que en la Tierra, pues tendrían que soportar pesos mucho menores en comparación. Y evidentemente, ningún animal astriano puede tener un tubo abierto que se extienda desde la boca hasta el ano, pues de tenerlo quedaría dividido en dos.

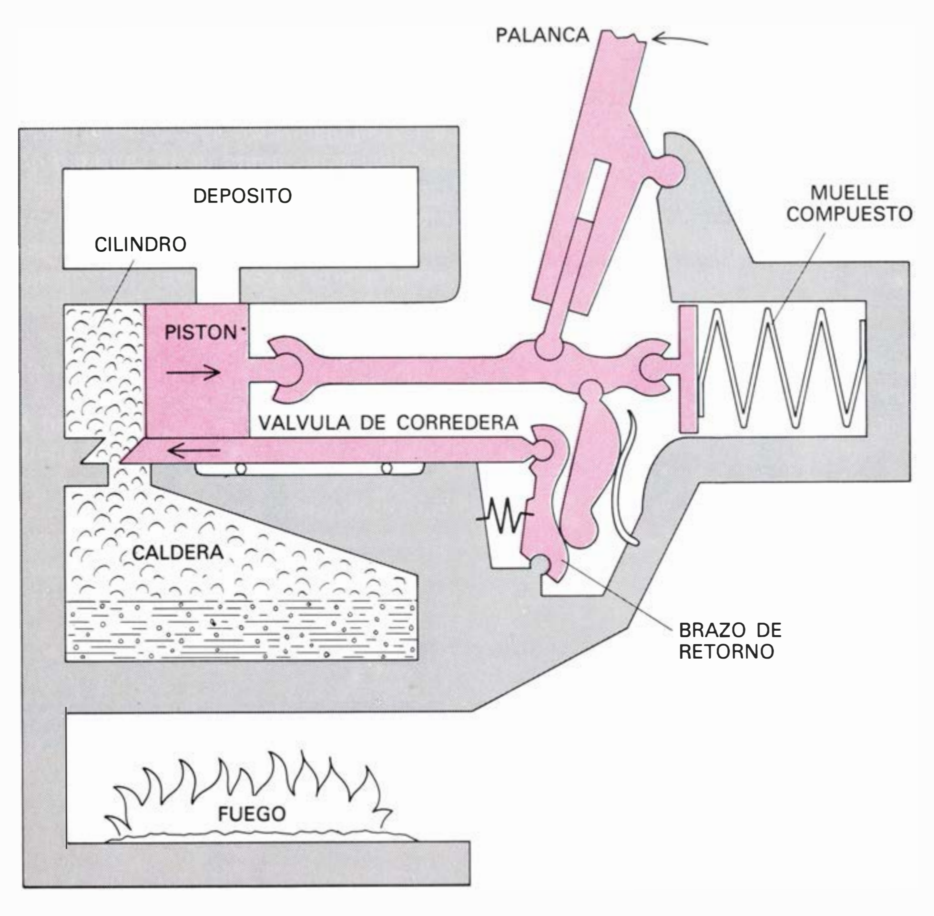
G. H. Whitrow, en el apéndice de su libro *The Structure and Evolution of the Universe* (Harper, 1959), arguye que la inteligencia no podría evolucionar en un espacio bidimensional, a causa de las grandes limitaciones que las dos dimensiones impondrían sobre las conexiones nerviosas. "En tres o más dimensiones", escribe allí, "cualquier número de neuronas pueden estar conectadas (unas con otras) por pares sin que se produzcan intersecciones de los enlaces, pero en dos dimensiones el número máximo de células para las que esto es posible es de solamente cuatro". Dewdney reduce a polvo semejante argumento, señalando que si se admite que las células nerviosas puedan disparar impulsos a través de los "puntos de cruce" se podrían formar redes nerviosas planas tan complejas como cualesquiera otras del esteriverso. Empero, las mentes de los

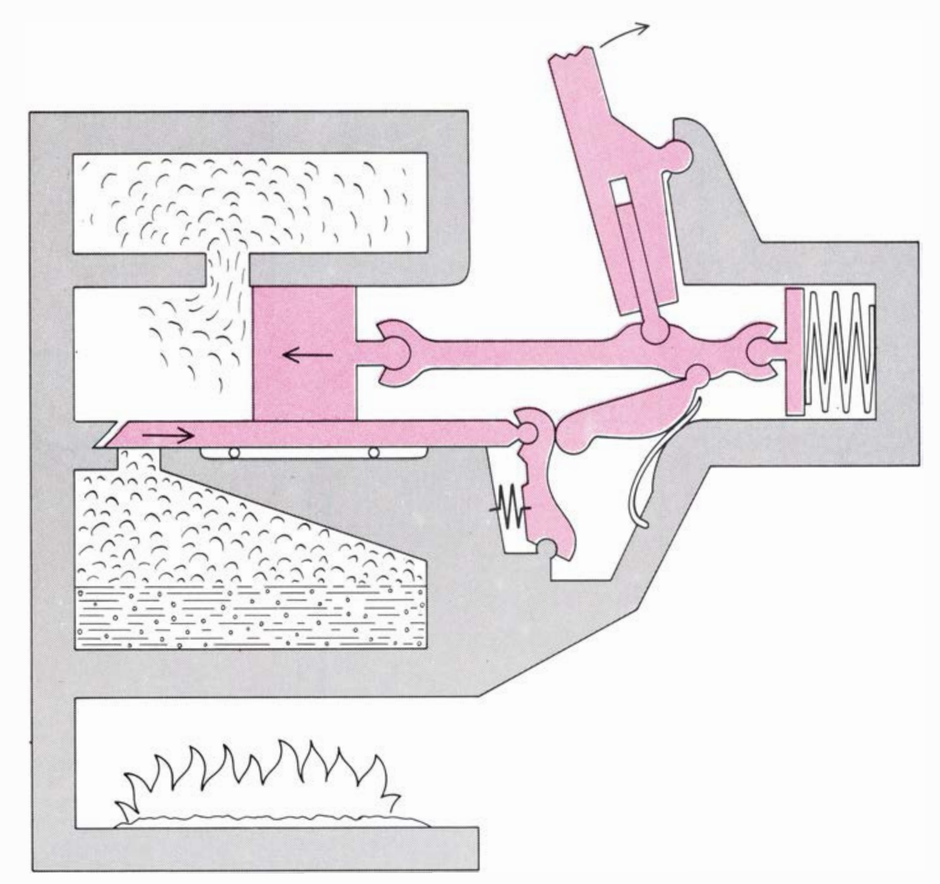


Elementos mecánicos planiversales fundamentales


Grifo astriano


Mecanismo de apertura de puertas astrianas





Máquina de vapor planiversal

seres planiversales tendrían un funcionamiento más lento que las esteriversales, pues en las redes bidimensionales los impulsos nerviosos encontrarían muchas más interrupciones. (Se dispone de resultados semejantes a éstos en la teoría de autómatas bidimensionales.)

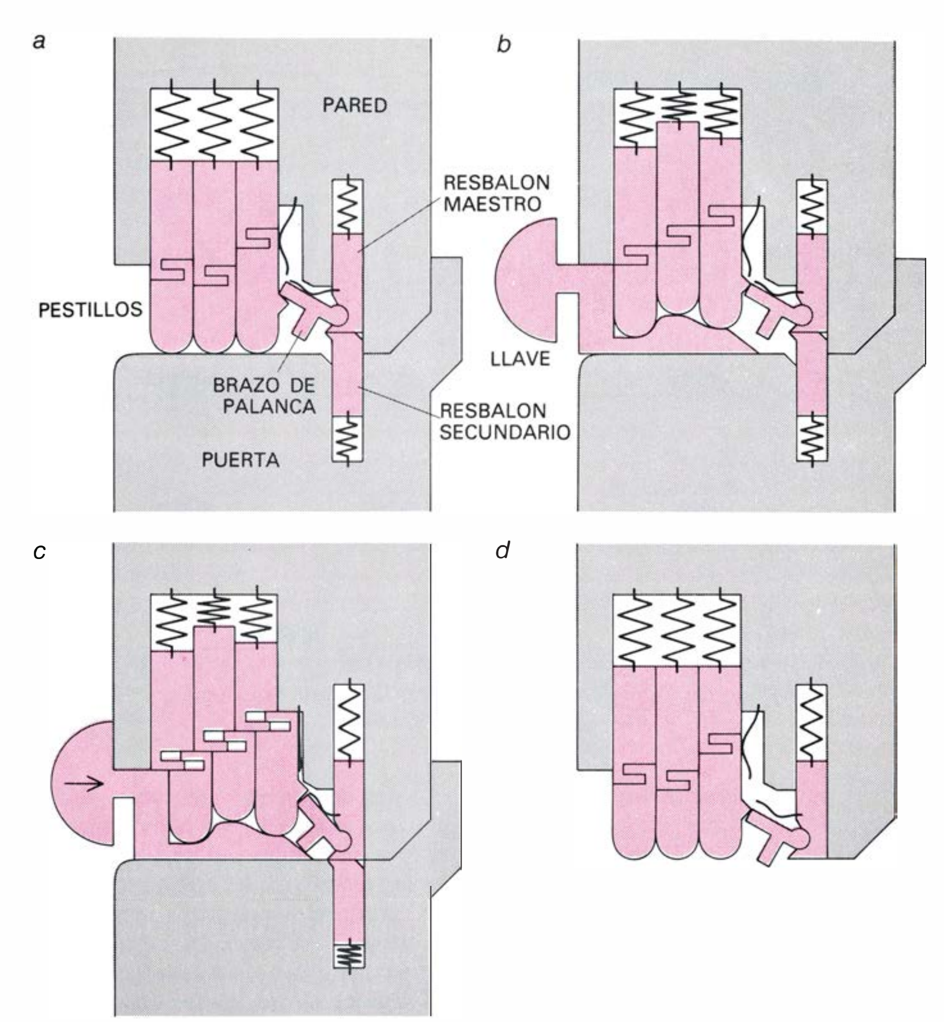
Dewdney dibuja con detalle la anatomía de una hembra de pez astriano, dotada de un saco de huevos no fecundados entre los dos músculos natatorios. El pez tiene esqueleto externo; su nutrición se consigue por circulación interior de vesículas alimenticias. Fijémonos en una célula aislada: el alimento penetra en ella a través de una membrana que solamente puede tener una abertura cada vez. Cuando la célula entra en contacto con otras células, como en los tejidos, puede tener más de un agujero por vez, pues las células circundantes pueden servir de soporte para mantenerla completa. Y como es obvio, podemos ver todos los órganos internos del pez o de cualquier otra forma de vida del planiverso, lo mismo que un animal tetradimensional podría observar todos nuestros órganos internos.

Dewdney adopta el convenio de Hinton, y, al igual que él, representa esquemáticamente a los pobladores de Astria, dibujándolos como triángulos con dos brazos y dos piernas. Sin embargo, los astrianos de Hinton siempre encaran en la misma dirección: los hombres hacia el este y, las mujeres, hacia el oeste. En ambos sexos los brazos están situados en el lado delantero; en el vértice superior del triángulo hay un único ojo, como se muestra en la ilustración superior de la página 104, a la izquierda. Los astrianos de Dewdney, en cambio, tienen simetría bilateral, y disponen de un brazo, una pierna y un ojo en cada costado según vemos en el centro de la misma ilustración. Por tanto, estos astrianos pueden ver en direcciones opuestas, lo mismo que los pájaros o los caballos terrestres. Como es obvio, la única forma de que un astriano adelante a otro es saltar por encima de él, o arrastrarse por encima o por debajo. En la parte derecha de la ilustración se ve la idea que me he formado de un posible monstruo de ojos saltones. Las extremidades de la criatura tanto pueden servir de brazos como de piernas, según de qué lado se vuelva; sus dos ojos le proporcionan visión binocular. Si tuvieran un solo ojo, los astrianos poseerían un mundo visual básicamente unidimensional, y con ello, una muy

restringida percepción de la realidad. Por otra parte, podrían distinguirse por el color las distintas partes de los objetos, y la ilusión de profundidad podría crearse por el enfoque de la lente del ojo.

La construcción de una casa, o encargarse de segar el césped, requiere en Astria menos trabajo que en la Tierra, pues la cantidad de material a manejar es considerablemente más pequeña. Pero como Dewdney hace notar, en un mundo bidimensional quedan todavía por resolver problemas formidables. "Admitiendo que la 'superficie' del planeta sea absolutamente esencial para sustentar los animales y plantas, es obvio que muy poco de su contorno podría perturbarse sin provocar la destrucción biológica del planeta. Por ejemplo, acá en la Tierra podemos construir una pequeña carretera que atraviese unas cuantas hectáreas de tierra cultivable, sin destruir más que un pequeño porcentaje de ella. Pero en Astria las carreteras destruirían *toda* la tierra cultivable sobre la que pasasen... Análogamente, las grandes ciudades agotarían rápidamente toda la campiña astriana. Al parecer, la única posibilidad viable para una sociedad tecnológica astriana sería la subterránea". En la ilustración inferior de la página 104 se muestra una casa subterránea típica, con sala de estar, dos dormitorios y desván. Para facilitar el paso por las habitaciones, el mobiliario es plegable, y se recoge en nichos, en el suelo.

Entre los muchos elementos mecánicos simples del mundo tridimensional que admiten análogos evidentes en Astria tenemos barras y varillas, palancas, planos inclinados, muelles, bisagras, cuerdas y cables. Es posible hacer rodar las ruedas sobre el suelo, pero no hay forma de hacerlas girar alrededor de un eje. Tornillos y tuercas son imposibles. También es imposible anudar cuerdas; por otra parte, éstas jamás pueden enredarse. Tubos y cañerías han de tener tabiques que mantengan los lados en su sitio, tabiques que es preciso ir abriendo (aunque nunca todos a la vez) para dejar pasar cuerpos por el tubo. Es sorprendente que pese a tan estrictas restricciones sería posible construir muchos dispositivos mecánicos planos que funcionarían perfectamente. En la ilustración central de la página 105 vemos un grifo diseñado por Dewdney. Para hacerlo funcionar se alza la manilla; se separará el obturador de la pared de la espita, dejando salir el agua. Al soltar la manilla, el muelle vuelve a cerrar la válvula.

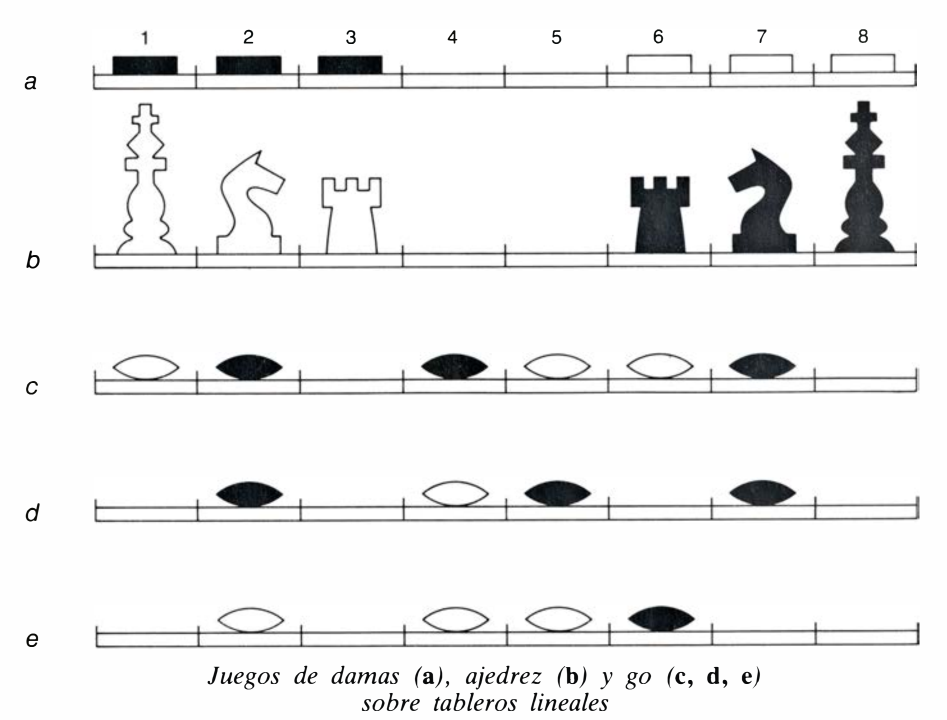


Cerradura y llave planiversales

El dispositivo que vemos en la ilustración inferior de la página 105 sirve para abrir o cerrar una puerta o una pared. Tirando hacia abajo de la palanca del lado derecho obligamos a la cuña situada al pie a desplazarse hacia la izquierda, permitiendo así que la puerta pivote hacia arriba (llevando consigo cuña y palancas) alrededor de una bisagra de la parte superior. Para abrir la puerta desde la izquierda se empuja hacia arriba la otra palanca. Moviendo la palanca en dirección apropiada podemos bajar la puerta desde cada uno de sus lados, y devolver la cuña a su lugar para estabilizar la pared. Lo mismo que el grifo, este dispositivo está dotado de bisagras planiversales permanentes: muñones circulares alojados en cavidades adecuadas, de donde no pueden desmontarse.

En la ilustración de la página opuesta vemos una máquina de vapor para el planiverso, que remeda en su funcionamiento a la esteriversal. El vapor a presión penetra en el cilindro del motor a través de una válvula de corredera, que forma una de las paredes (*arriba*). La presión del vapor provoca el desplazamiento del pistón hacia la derecha, hasta que el vapor logra escapar a una cámara de almacenamiento situada sobre él. La consiguiente pérdida de presión permite que el resorte múltiple situado a la derecha del émbolo impela al pistón hasta el principio de su carrera (*abajo*). Conforme el vapor escapa hacia la cámara se cierra la válvula de corredera, pero al retornar el pistón la válvula vuelve a abrirse, impulsada hacia la derecha por un brazo cargado por un muelle.

La ilustración de esta página exhibe el ingenioso mecanismo ideado por Dewdney para cerrar con llave una puerta. Este cerrojo planiversal está formado por tres pestillos ranurados (*a*) que al insertarse la llave quedan alineados (*b*), de manera que sus mitades inferiores restan solidarias, moviéndose como un todo al empujar con la llave (*c*). El avance de la llave se transmite mediante un brazo de palanca hasta el resbalón maestro, que empuja hacia abajo a un resbalón secundario hasta dejar libre a la puerta para



*Juegos de damas (**a**), ajedrez (**b**) y go (**c, d, e**) sobre tableros lineales*

girar hacia la derecha (*d*). La barra situada sobre el brazo de palanca y el pico consiguen que la cerradura sea difícil de abrir con ganzúa. Mediante resortes de ballestas (simples o múltiples) se hacen retornar todas las piezas de la cerradura a su posición inicial cuando se abre la puerta y se extrae la llave. Cuando la puerta se cierra, golpea contra la barra situada sobre el brazo de palanca, haciendo retornar también esta pieza a su posición primitiva. Esta cerradura plana podría emplearse también en el esteriverso: bastaría insertar la llave en la cerradura, sin necesidad de darle la vuelta. Al respecto escribe Dewdney: "Resulta gracioso pensar que las insólitas exigencias que el entorno bidimensional impone a los diseños pudieran obligarnos a enfocar los mecanismos desde un ángulo tan diferente que llegasen a surgir soluciones totalmente nuevas a problemas ya viejos. Los diseños resultantes, de ser prácticos en el esteriverso, invariablemente economizan espacio".

Quedan por resolver en el planiverso miles de apasionantes problemas. ¿Existirá forma, se pregunta Dewdney, de diseñar un motor de muelles planos o tiras de goma capaz de almacenar energía? ¿Cuáles serían los diseños planiversales óptimos de relojes, teléfonos, libros, máquinas de escribir, coches, ascensores, u ordenadores? ¿Habrá máquinas que exijan sustituir el sistema rueda-eje? ¿Habrá alguna que exija energía eléctrica?

Tratar de inventar máquinas para lo que Dewdney describe como "un universo a un tiempo tan parecido y a la vez tan extrañamente distinto del nuestro" produce un curioso placer. Con sus propias palabras, "a partir de un pequeño número de hipótesis parecen desplegarse tantos fenómenos que nos dan la sensación de que este mundo bidimensional tiene existencia propia e independiente. Nos descubrimos a nosotros mismos hablando, sin darnos cuenta, *del* planiverso, como antagónico de *un* planiverso... (Para) quienes se embeben decididamente en él hay una especie de raro disfrute, como (el de) un explorador al pisar nuevas tierras, cuyas percepciones tienen papel fundamental al examinar el paisaje que saluda sus miradas".

Ciertos aspectos filosóficos de esta exploración distan mucho de ser triviales. Al construir un planiverso vemos inmediatamente que será imposible erigirlo sin introducir una multitud de axiomas, los que Leibniz llamó elementos "compossibile" de cualquier posible mundo, que permiten crear una estructura coherente desde el punto de vista lógico. Empero, como hace notar Dewdney, en nuestro universo la ciencia está basada fundamentalmente en observaciones y experimentos, y no es nada fácil descubrir los axiomas subyacentes. Al construir un planiverso no tenemos nada que observar. Sólo podemos realizar los que Einstein llamaba *gedanken Experimente,* experimentos con el pensamiento, concernientes a lo que podríamos observar si pudiéramos llevarlos a cabo. Dewdney lo expresa así: "la pérdida del experimentador es la ganancia del teórico".

A todos cuantos deseen participar en esta deliciosa empresa (o simplemente, saber más al respecto) les insisto en que se hagan cuanto antes con la monografía de Dewdney, quien me comunica que la enviará a portes pagados (contra un envío de 5 dólares en Estados Unidos y Canadá, y el recargo postal pertinente para los demás países). Los cheques o giros postales deben ir dirigidos a Dewdney, al Department of Computer Science, University of Western Ontario, London, Ont., Canada N6A 5B9. De llegar a comprometerse suficientes personas en la construcción de un planiverso, Dewdney estaría dispuesto a ir publicando anualmente sus contribuciones.

Podría prepararse una maravillosa muestra de máquinas planiversales, capaces de funcionar, recortadas en cartón o chapa metálica, que conviene montar sobre una superficie inclinada con objeto de simular la gravitación planiversal. Podríamos concebir también preciosas muestras de paisajes, ciudades y casas planiversales. Como vemos, lo que ha hecho Dewdney es inaugurar un nuevo juego, que tanto requiere saber de ciencias como de matemáticas, a saber, la exploración de un mundo tan vasto como fantástico y del que por el momento apenas sabemos nada.

Se me ocurre que aunque los astrianos fueran capaces de entender los juegos de tablero bidimensionales serían en ellos tan desmañados como nosotros lo somos en los tridimensionales. Más me los imagino, por tanto, enfrentados en diversos juegos lineales sobre el tablero análogo de nuestro damero de 8 por 8. En la ilustración de esta página presentamos varios juegos de este tipo. En la parte *a* vemos el comienzo de una partida de damas. Las piezas solamente pueden avanzar, haciéndolo a razón de una casilla por vez; los saltos son obligatorios. La versión lineal del juego es equivalente a una partida de damas ordinaria desarrollada exclusivamente sobre la diagonal principal de un tablero normal. Es fácil ver que el segundo jugador puede vencer siempre si juega racionalmente, y que en la versión "misère" o "a traición", es el primer jugador quien vence con igual facilidad. Las partidas de damas en tableros lineales van haciéndose progresivamente más difíciles de analizar conforme se usan table-

ros más y más largos. Por ejemplo, ¿qué jugador vencería en la versión normal de la partida de damas con tablero lineal de 11 casillas, cuando en la posición inicial cada uno de ellos dispone de cuatro piezas situadas en las cuatro primeras casillas de su extremo del tablero?

La parte *b* de la ilustración muestra un equivalente astriano del ajedrez, muy entretenido. En los tableros lineales, los alfiles carecen de papel, y las reinas son indistinguibles de las torres; las piezas, por tanto, se reducen a reyes, caballos y torres. La única modificación de reglas necesaria es que los caballos saltan dos cuadros en cada dirección, pudiendo saltar sobre cualquier pieza intermedia del mismo o distinto color. Si el juego se desarrolla racionalmente, ¿conseguirá ganar siempre alguno de los dos bandos, o se podrá terminar en tablas? La cuestión resulta sorprendentemente difícil de contestar.

El *go*, jugado sobre el mismo tablero lineal, tampoco es nada trivial. La versión que voy a describir fue inventada hace 10 años por James Marston Henle, un matemático hoy perteneciente al Smith College. Bautizado "pinch" por Henle, lo publicamos aquí por primera vez.

En una partida de "pinch" los jugadores van por turnos colocando piedrecitas blancas y negras en las casillas de un tablero lineal. Siempre que las piezas de un bando rodean a las del otro, se retiran las piezas rodeadas del tablero. Por ejemplo, los dos conjuntos de piedras de la parte *c* de la ilustración están rodeados. Para jugar al pinch deben respetarse las siguientes reglas:

Regla 1: No puede colocarse ninguna piedra en ninguna casilla donde quede rodeada, excepto si tal jugada sirve para rodear una colección de piezas contrarias. Por tanto, en la situación que vemos en la parte *d* de la ilustración, las blancas no pueden jugar en las casillas 1, 3 ni 8, pero sí pueden jugar en la casilla 6, porque esta jugada sirve para rodear el cuadro número 5.

Regla 2: No se puede colocar ninguna pieza en casillas de donde acaben de ser eliminadas piedras en la última jugada, cuando el objetivo de la jugada sea eliminar alguna otra pieza contraria. Los jugadores deben antes esperar al menos un turno. Por ejemplo, supongamos que en la parte *e* de la ilustración las negras ocupen la casilla 3 y eliminen las piezas blancas de las casillas 4 y 5. Las blancas no pueden jugar en la casilla 4 (y rodear así la casilla 3) en la jugada siguiente, aunque sí podrían hacerlo en cualquier momento posterior. Sin embargo, sí pueden ocupar la casilla 5, pues aunque de esa casilla acaba de eliminarse una piedra, la jugada de las blancas no sirve para rodear nada. La regla está proyectada para reducir el número de empates por ahogo, lo mismo que la regla análoga del *go*.

Con dos casillas es evidente que el segundo jugador tiene asegurada su victoria. Los juegos con tres y cuatro casillas dan al primer jugador fáciles victorias sin más que tomar al salir la casilla central (en el juego de tres) o una de las dos centrales (en el caso de cuatro). Con cinco cuadros es seguro que vence el segundo jugador, mientras que con seis o siete casillas vuelve a ser el primero quien consigue la victoria. Las partidas con ocho escaques alcanzan ya tal grado de complejidad que el juego se torna apasionante. La fortuna va cambiando rápidamente de bando, y, en la mayoría de los casos, el jugador que dispone de estrategia vencedora tiene solamente una jugada que le dé la victoria.

El mes que viene revelaré qué jugador lleva las de ganar en el pinch de ocho casillas, y responderé las cuestiones sobre los otros tres juegos lineales.

El problema del mes pasado consistía en determinar si los tres modelos de grupos de orden 4 eran ejemplos de grupos cíclicos o del grupo de Klein. Los tres resultan ser grupo de Klein. Un criterio sencillo para determinar el carácter de un grupo de orden 4 consiste en comprobar si cada operación perteneciente al grupo es inversa de sí misma. De ser así, el grupo es el grupo de Klein de orden 4.

Los dos grupos de orden 4 nos proporcionan el ejemplo más sencillo que muestra cómo los grupos simples, pese a ser los bloques con que se construyen todos los grupos finitos, no se comportan de forma análoga a los números primos. Ambos grupos de orden 4 contienen el mismo sistema de subgrupos simples: el grupo identidad, y tres subgrupos de orden 2, formados cada uno de ellos por la identidad y por una de las tres operaciones restantes. Empero, pese a tener los mismos "divisores", los dos grupos de orden 4 son distintos. Y mientras que el producto de un conjunto de números primos da siempre un único número, un mismo conjunto de grupos simples puede combinarse de distintas formas, originando más de un grupo compuesto.

Taller y laboratorio

Aprovechando cómo actúan las fuerzas, el judo logra que los débiles compitan con los fuertes

Jearl Walker

El judo y el aikido son artes marciales para las cuales se requiere una comprensión intuitiva de las leyes físicas relativas a las fuerzas, momentos de giro, equilibrio y movimientos de rotación. Este mes voy a examinar algunas de las llaves fundamentales de estas dos formas de lucha. No me es posible expresar exactamente la habilidad y desenvoltura que exige cada una de estas llaves; no obstante, sí es posible separarlas en componentes que puedan examinarse en términos de la mecánica clásica. Las experiencias que voy a describir necesitan la puesta en práctica real de las llaves, pero ello sólo deberá hacerse bajo la vigilancia de un experto, ya que ambos tipos de lucha pueden resultar peligrosos para el lector y su contrincante.

En el judo, el propósito principal es hacer perder el equilibrio al contricante. El arte reside en anticiparse a sus movimientos y en la coordinación de las respuestas propias, con la idea de evitar forzarle a una resistencia firme que opondría su fuerza física a la nuestra. Siempre que eluda las pruebas de fuerza, todo practicante de judo de corta estatura y bien adiestrado tendrá una ventaja muy marcada frente a contrincantes de mayor estatura pero inexpertos.

Probablemente el mejor ejemplo de esta ventaja lo encontremos en la proyección de cadera básica, cuya mayor eficacia se da frente a un adversario de mayor estatura y más lento. En un encuentro de judo normal nos enfrentamos al contrario asiéndole con las manos las solapas u hombros de su vestimenta. Para ejecutar esta proyección se adelanta el pie derecho hasta un punto situado entre los pies del contrario, tirando de él hacia abajo y hacia nuestra derecha. La llave surte efecto si se le ha sorprendido en el preciso momento en que adelantaba el pie derecho. En ese instante, seguirá manteniéndose estable si tiramos de él de frente hacia nosotros; pero, a causa de la posición de sus pies, lo será considerablemente menos si tiramos de él hacia nuestra derecha.

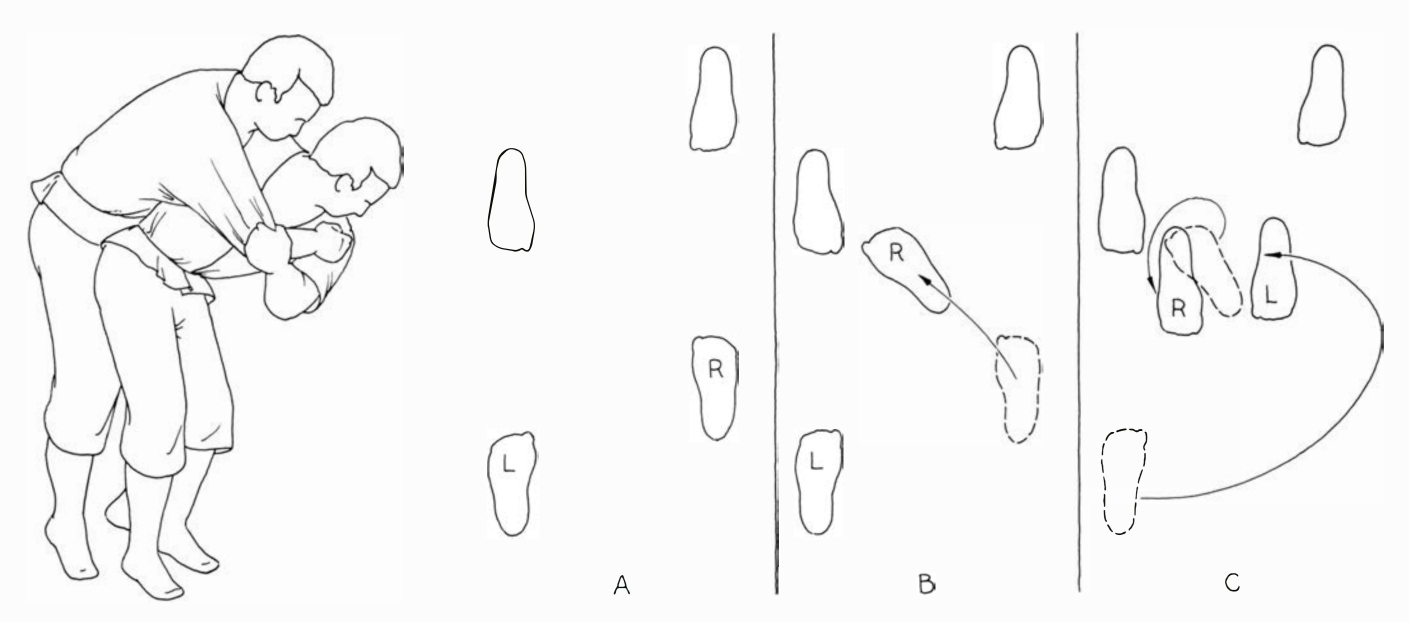
A la vez que se adelanta el pie propio hay que encorvar el cuerpo hacia adelante, de forma que la cabeza nos quede a la altura del hombro del contrario. Después, y muy rápidamente, se gira hacia atrás la cadera izquierda al tiempo que se tira de él hacia nuestra cadera derecha. Aquí debe producirse el primer contacto de ambos cuerpos durante el movimiento. Si seguimos tirando de él y girando hacia atrás la cadera izquierda hasta que nos encontremos mirando en la misma dirección que el contrario, a éste le habremos hecho girar en torno a nuestra cadera derecha, proyectándolo sobre la colchoneta.

Como al contrario no se le debe lastimar, durante la caída seguimos asiéndole de la vestimenta, de suerte que aterrice sobre su lado izquierdo y pueda dar un palmetazo a la colchoneta con la mano izquierda durante el impacto. El efecto del palmetazo es repartir la fuerza de impacto sobre una superficie mayor, con lo que la presión sobre sus costillas no será suficiente para lastimarlo. Gran parte del entrenamiento de judo para principiantes se refiere a cómo sincronizar el palmetazo para que coincida con el impacto. La única vez que resulté lastimado practicando este deporte fue cuando no conseguí dar el palmetazo correctamente.

En la proyección de cadera son fundamentales la coordinación y una ejecución suelta y fluida, pero también es necesario comprender su mecánica, en especial en lo referente a los momentos de giro y al centro de masa. El centro de masa del contrario es el centro geométrico de su distribución de masa, y puede considerarse como el punto donde la fuerza gravitatoria actúa sobre el conjunto de su cuerpo, razón por la cual a veces se llama centro de gravedad. El oponente se encuentra en equilibrio en tanto que su centro de masa se mantenga encima de la superficie de apoyo delimitada por sus pies. Cuando permanece erguido en postura normal, tiene su centro de masa situado aproximadamente entre la espina dorsal y el ombligo; por tanto, estará en equilibrio hasta que le forcemos o le engañemos para que desplace su centro de gravedad, o para que pierda parte de la superficie de apoyo.

Supongamos que durante una llave conseguimos desplazar el centro de masa del contrario delante de sus pies. Entonces, aun cuando ya no apliquemos la llave, la atracción gravitatoria actuante en su centro de masa creará un momento de giro capaz de derribarlo. Para calcular un momento de giro hay que multiplicar dos cantidades: la fuerza actuante que produce la rotación y el brazo de palanca entre el centro de giro y la fuerza. El brazo de palanca es la distancia en perpendicular desde el centro de giro hasta la línea recta que contiene a la fuerza. Cuando hayamos hecho perder la estabilidad al contrario, la fuerza que puede hacerlo girar en torno de sus pies, llevándolo a la colchoneta, es la atracción gravitatoria. Esta fuerza, que no es sino su peso, la representaré mediante un vector que parte del centro de masa y señala hacia abajo. Así pues, el brazo de palanca será la distancia horizontal desde el centro de giro situado en sus pies y una prolongación de la recta vertical que contiene al vector peso. El momento de giro actuante sobre un contrario desestabilizado es el producto de su peso por el brazo de palanca. Cuando el contrario está erguido, el brazo de palanca correspondiente a su vector peso es nulo, por lo que el momento de giro será también nulo. Cuando se le sorprende con el centro de masa en una posición adelantada respecto a sus pies, el brazo de palanca ya no es cero y el momento de giro resultante provoca su rotación. Cuanto mayor sea el brazo de palanca (o sea, cuanto más inclinado esté), tanto mayor será el momento de giro. En el judo, uno de los objetivos es engañar al contrario para llevarlo a una posición inestable con tanta rapidez que no le sea posible recuperarse y, una vez conseguido esto, continuar la ejecución de la llave aplicándole otro momento de giro tal que acabe en la colchoneta antes de que intente siquiera recobrar el equilibrio.

En la proyección de cadera se comienza tirando de la vestimenta del contrario con el fin de hacerle perder la estabilidad. Si tirásemos directamente hacia nosotros, no sería fácil conseguir esta inestabilidad, ya que su centro de masa se trasladaría por encima de su pie adelantado, pero entonces él podría mantener el equilibrio doblando la rodilla correspondiente. En estas condiciones, para hacerle perder estabilidad tendríamos que desplazar su centro



Proyección de cadera básica del judo

Juego de pies en la proyección de cadera

de masa una distancia relativamente grande hasta que estuviese fuera de su pie adelantado, movimiento que requeriría un esfuerzo de tracción vigoroso y prolongado, que él podría contrarrestar rápidamente.

Una manera más sencilla de hacer perder estabilidad al contrario es tirar de él hacia nuestra derecha, ya que en dicha dirección su centro de masa debe recorrer sólo una distancia relativamente corta para alcanzar una vertical que no pase por la superficie de apoyo. Probablemente no pueda contrarrestar un tirón de este tipo antes de que aparezca la inestabilidad y, a partir de ese instante, será incapaz de neutralizar la rotación continuada que implica una proyección de cadera.

Nuestro tirón posee un objeto adicional, que es doblar el cuerpo del contrario de forma que su centro de masa se le adelante hasta el ombligo, o justo fuera del cuerpo. Esta nueva posición nos facilita después el hacer girar el cuerpo del contrario sobre nuestra cadera derecha. Efectuado el contacto entre ambos cuerpos, en nuestra cadera se establece un nuevo centro de giro y la fuerza de tracción crea un nuevo momento de giro sobre el contrario, el cual provoca la rotación de éste sobre nuestra cadera. Como antes, el momento de giro se calcula multiplicando la fuerza que actúa sobre el contrario por el brazo de palanca que va desde la recta de la fuerza al centro de giro. Esta vez la fuerza es el tirón y el centro de giro es la cadera propia. Resulta así que, en la proyección de cadera, se originan dos momentos de giro que actúan sobre el contrario, uno de ellos debido a su propio peso y postura inestable y el otro debido directamente a la tracción que le ejercemos. La llave comienza con el primer momento de giro con el fin de que podamos establecer el segundo sin que él ofrezca resistencia.

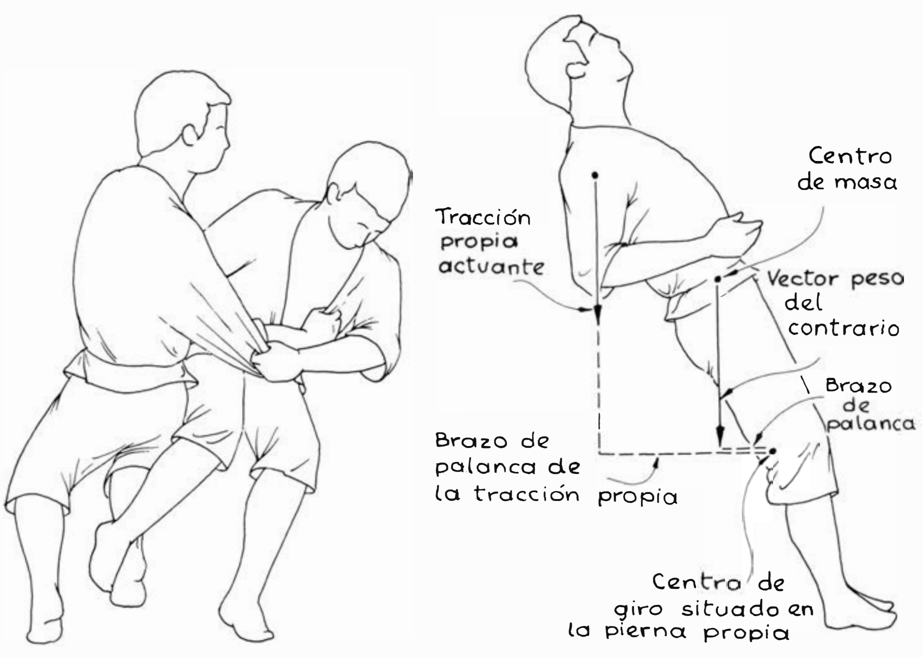
Supongamos que no doblamos el cuerpo del contrario y que, por tanto, su centro de masa no queda fuera del ombligo. Entonces, cuando intentáramos hacer girar su cuerpo sobre nuestra cadera, habría en realidad un momento de giro producido por su peso que contrarrestaría el momento de giro



Pérdida del equilibrio a causa del peso

Fuerzas actuantes en la proyección

Proyección de cadera incorrecta



«Siega externa mayor» *Fuerzas que intervienen en la proyección*

producido por nuestro tirón. Supongamos que siga erguido cuando establecemos contacto con su cuerpo y tratamos de hacerle girar sobre nuestra cadera. El punto de apoyo de la rotación es la cadera propia; vamos a utilizarla para determinar los brazos de palanca de los dos momentos de giro que en ese instante actúan sobre el contrario.

Uno de los momentos de giro es el producto de nuestro esfuerzo de tracción por el brazo de palanca que va desde nuestra cadera a la recta que contiene al vector que representa dicho esfuerzo de tracción. El otro momento de giro es el producto del peso del contrario por el brazo de palanca que va desde nuestra cadera a la vertical que pasa por su centro de masa. Si el contrario permanece erguido, el momento de giro producido por su peso se opone al momento de giro que nosotros le aplicamos, ya que tiende a hacerlo girar sobre nuestra cadera en sentido contrario. Entonces, para finalizar la llave, debemos vencer el momento de giro debido a su peso, pero el tiempo necesario para ello destruirá la ventaja de la sorpresa. Y lo que es más, deberemos entrar en una prueba de fuerza con el contrario.

Cuando la proyección de cadera se ejecuta correctamente, el cuerpo del contrario se dobla formando una curva y su centro de masa se traslada delante de su ombligo, de manera que el brazo de palanca asociado a su peso disminuye o desaparece. Con ello decrece el momento de giro debido a su peso y podemos hacer girar su cuerpo sobre nuestra cadera derecha con relativa facilidad. Esta llave es más eficaz contra un contrincante más alto, dado que podemos tirar de él hacia abajo para encorvarlo adecuadamente con mayor facilidad que si se tratara de un contrario de estatura igual o menor que la nuestra. Además, es mucho más sencillo introducir la cadera bajo un contrario de mayor estatura y también será mayor el brazo de palanca del tirón sobre su vestimenta, produciéndose así un momento de giro mayor para llevarlo a nuestra cadera.

La llave llamada "siega externa mayor" (en japonés *osotogari*) se entiende algo mejor en base a los movimientos de rotación. Cuando el contrario retrasa el pie izquierdo, nosotros adelantamos el pie izquierdo colocándolo por la parte exterior de su pie derecho a la vez que tiramos hacia abajo de su vestimenta, de forma que obligamos a que su peso descanse sobre dicho pie. Además, el tirón debe efectuarse en dirección a su retaguardia y hacia su derecha para que se curve hacia atrás. Así ya le habremos llevado a una posición inestable, puesto que habremos desplazado su centro de masa hacia la derecha de su retaguardia separándolo de la superficie de apoyo delimitada por los pies. No le es posible escabullirse deslizando el pie hacia la derecha y recobrando el equilibrio, puesto que ya le hemos forzado hacia abajo. Su inestabilidad resulta del momento de giro creado por el peso respecto a un centro de giro situado en sus pies, primordialmente el derecho. El brazo de palanca va desde este centro de giro a la recta definida por el vector de su peso. Situado en esta posición ya lo tenemos preparado para aplicarle la segunda parte de la llave,



Barrido de tobillo



Ejecución de una presa de brazo en cruz

en la que eliminaremos aún más su superficie de apoyo y le aplicaremos un segundo momento de giro que lo llevará rápidamente a la colchoneta.

Continuamos la llave adelantando el pie derecho por detrás y alrededor del pie derecho del contrario. Seguidamente, llevamos la cadera y pierna derechas hacia nuestra retaguardia a la vez que forzamos su costado derecho hacia la derecha de su retaguardia, dejándole prácticamente sin base de apoyo. Los dos momentos de giro le harán girar en torno a un punto situado en nuestra pierna derecha y, aun cuando no siguiéramos empujándolo después de barrerle la pierna, él girará en torno de nuestra pierna dado que la gravedad atraerá su centro de masa; en realidad, nuestro empujón hacia abajo sirve para acelerar su caída. En esta llave los dos momentos de giro se complementan entre sí.

En el "barrido de tobillo" (*okuriashi bari*) se elimina el apoyo del contrario sobre una de sus piernas de una manera similar. En el momento en que éste se dispone a descansar su peso sobre el pie derecho en el transcurso de un paso adelante o atrás, llevamos nuestro pie izquierdo a un punto de dicha pierna situado justo encima del tobillo. Simultáneamente, tiramos de su vestimenta en la dirección de su desplazamiento original. Supongamos que estaba moviéndose hacia adelante. Entonces tiramos hacia adelante (y así no encontraremos resistencia) a la vez que barremos su pie derecho hacia su pie izquierdo. Así, aun cuando consiga mantener el pie izquierdo sobre la colchoneta, su superficie de apoyo se habrá reducido notablemente y la habremos desplazado de por debajo de su centro de masa, con lo que su peso producirá un mo-



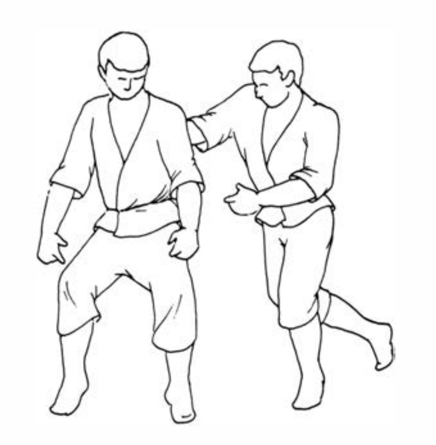


Movimiento de aikido para eludir un intento de apresamiento de las muñecas desde atrás





Defensa contra un golpe de canto de mano





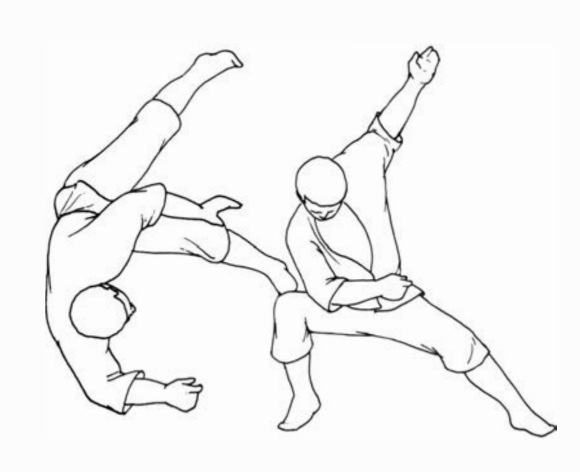
Defensa contra un intento de apresamiento de los brazos desde atrás

mento de giro que lo llevará a la colchoneta. Si además inclinamos el cuerpo a la vez que mantenemos el tirón sobre su vestimenta, habremos desarrollado otro momento de giro que lo hará girar hacia la colchoneta. El centro de giro es aquí su pie izquierdo y vuelve a ocurrir que los dos momentos de giro se complementan entre sí.

En las clases superiores de judo se enseñan procedimientos para poner fuera de combate a un contrincante sobre la colchoneta ("técnicas en el suelo"). Estas "inmovilizaciones" implican en su mayoría el apresamiento del contrario sobre la colchoneta, con el peso propio aplicado de tal forma que le sea imposible darse la vuelta o levantarse, aun cuando sea más fuerte. Por ejemplo, en la "presa de brazo en cruz" (*udehishigi-jujigatame*) la posición propia es tal que parte de nuestro peso actúa sobre el torso del contrario, que se encuentra en el suelo y en tal situación que no sólo le es imposible levantarse, sino que además es probable que ni siquiera intente moverse por temor a partirse el brazo.







El movimiento se origina estando nosotros a horcajadas junto al costado del contrario, quien se encuentra postrado boca arriba. Cuando éste levanta el brazo izquierdo con el ánimo de prevenir nuestro ataque, le sujetamos la muñeca con ambas manos, nos dejamos caer a su lado, proyectamos la pierna derecha sobre su cuello y con la rodilla izquierda alzada lanzamos el tobillo izquierdo contra su costado. Así tendremos su brazo izquierdo con el codo hacia abajo apresado entre nuestras piernas. Entonces, incluso un leve empuje hacia abajo sobre su muñeca creará un enorme momento de giro sobre su brazo en torno al punto en que éste se cruza con nuestra pierna derecha, punto que es el centro de giro. No podrá sentarse ya que nuestro peso creará sobre él un momento de giro abrumador cuando intente hacer rotar el tronco en torno a un centro de giro situado a la altura de sus caderas. Tampoco podrá liberar su brazo izquierdo aun cuando sea considerablemente más fuerte que nosotros. Podría tratar de contrarrestar la acción del momento de giro que le aplicamos a su brazo utilizando los músculos de los hombros, pero éstos actuarían sobre un punto de su brazo muy cercano al centro de giro, con lo que dispondrían de un brazo de palanca muy corto. Como en la mayoría de las técnicas de judo, una persona adiestrada para crear los momentos de giro adecuados sobre un contrincante goza de una ventaja enorme, por más que aquél sea mucho más fuerte.

El aikido es una forma de arte marcial relativamente moderna en la que se reúnen técnicas tomadas de muchas de las otras. Su rasgo más característico es la estricta norma de evitar toda lesión al contrario. Por ello, más que un deporte de competición es una forma de autodefensa en la que no intervienen técnicas que puedan considerarse ataques. Por mi parte opino que es la más difícil de aprender de todas las artes marciales, pues sus exigencias de destreza, desenvoltura y coordinación rivalizan con las del ballet.

En el aikido se emplean en gran parte los mismos principios de mecánica que se encuentran en el judo. Supongamos que el contrario nos sujeta las muñecas desde atrás. En uno de los movimientos del aikido hacemos descender suavemente nuestro cuerpo a la vez que elevamos las muñecas por encima de la cabeza y hacia adelante. El contrario no habrá soltado nuestras muñecas pero al descender le habremos desplazado e inclinado ligeramente hacia adelante. Por consiguiente, su posición será inestable ya que su centro de masa se encontrará algo adelantado respecto a sus pies. Rápidamente, entonces, hacemos retroceder la pierna derecha y nos dejamos caer sobre la rodilla derecha, adelantando el torso para formar un arco hacia abajo con éste y los brazos. Como en la primera parte del movimiento se indujo al contrario a sujetar con fuerza nuestras muñecas, ahora le proyectamos hacia adelante por encima de nuestro cuerpo.

Como ocurre casi siempre en el aikido, es nuestro contrario quien en realidad se proyecta a sí mismo. El no puede impedir nuestro movimiento hacia adelante a causa de la postura inestable en que le colocamos al principio y, aunque su peso sea mayor, no podrá detenerlo tirando hacia abajo de nuestras muñecas ya que, en una postura como la suya, sólo podrá tirar a lo largo de nuestros brazos en cuyo caso el momento de giro de su tirón será nulo al no existir brazo de palanca. Recuérdese que el brazo de palanca es la distancia perpendicular desde el centro de giro (nuestro hombro en este caso) hasta la recta sobre la que se encuentra la fuerza y ésta, para el tirón ejercido por nuestro contrario, pasa por nuestro hombro y por tanto carece de brazo de palanca. Así, aunque se trate de una persona pesada o vigorosa, no podrá hacer girar nuestros brazos una vez los hayamos colocado correctamente por encima de la cabeza.

Muchas técnicas del aikido se basan en la desviación de una fuerza dirigida directamente hacia nosotros. Supongamos que el atacante nos lanza un puñetazo al rostro. Para detener un puñetazo de frente se requiere resistir una gran fuerza de impacto, quizá superior a 3000 newton, que probablemente produzca rotura de huesos. Un procedimiento más prudente es desviar el golpe ya que, aunque para pararlo de frente sea necesaria una fuerza elevada, para desviarlo basta una fuerza reducida. Así, una fuerza de 10 newton puede ser suficiente para desviar un centímetro un golpe directo.

Si bien la mayoría de los partidarios de los estilos de lucha occidentales consideran que el ataque es una ventaja, en el aikido el atacante se encuentra en inferioridad manifiesta a causa del impulso de su propio golpe, puesto que este impulso se emplea para derribarlo. Supongamos que el atacante comienza adelantándose con el pie derecho y lanzándonos un golpe a la cara con el canto de la mano derecha (ataque típico tanto de los estilos de lucha occidentales como del kárate). Entonces, retrasamos el pie derecho a la vez que evadimos el golpe con el brazo izquierdo. Con esto buscamos desviar el golpe, no pararlo y ni siquiera frenarlo, ya que cualquiera de ambas cosas requeriría de nosotros una fuerza considerable. Durante este movimiento se guía el brazo derecho del atacante hacia abajo hasta asirlo con nuestra mano derecha y, mientras seguimos sin oponernos a su impulso hacia adelante, tiramos de él en redondo aprovechando el movimiento de rotación que hemos iniciado al retirar hacia atrás nuestra pierna izquierda. Al mantener adelantado el pie derecho, nuestro atacante se encontraba en una posición relativamente estable frente a una tracción directamente hacia adelante, pero será muy inestable frente a una tracción ejercida hacia adelante y hacia su izquierda, ya que en esta dirección no tendremos que desplazar mucho su centro de masa para que éste se haga inestable y caiga. Por tanto, a la vez que seguimos girando, tiramos de él en tal dirección. Ahora lo tendremos en una situación gravemente desfavorable por dos motivos. Primero, ha adquirido un movimiento hacia adelante cuya detención exigiría una fuerza considerable, incluso de él mismo. Segundo, nuestro tirón y su movimiento están haciendo salir su centro de masa fuera de su base de apoyo.

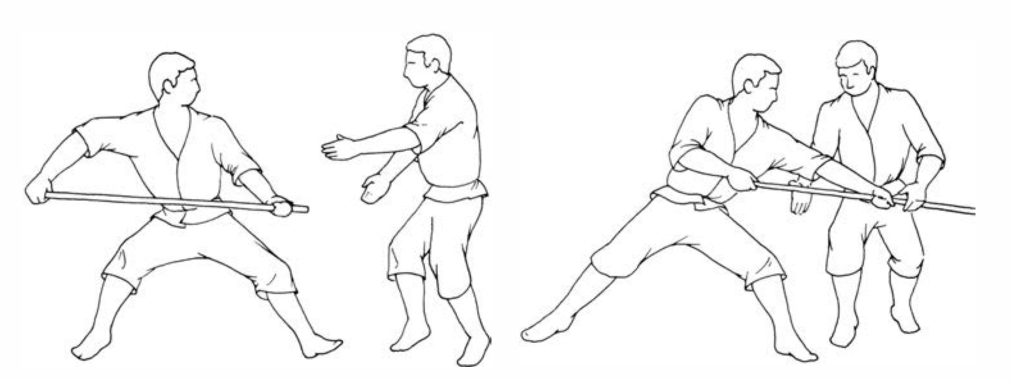
Para terminar la proyección hacemos girar hacia abajo el brazo derecho de nuestro atacante a la vez que retrocedemos hacia la izquierda de nuestra retaguardia, le retorcemos la muñeca de arriba abajo y le doblamos la mano sobre ella. En este punto ya le será imposible impedir ser derribado, pues habrá perdido el equilibrio y será totalmente incapaz de detener su propio movimiento. Además, tampoco podrá desasirse de nuestro agarrón porque le hemos doblado el brazo a la altura de la muñeca y, por muy fuertes que sean sus brazos, no podrá evitar el momento de giro que le hemos creado al empujarle la mano en torno a su propia muñeca. O sea, que caerá sobre la colchoneta.

¿Cómo se hace frente a un golpe a la cabeza de esta naturaleza en el kárate? En el *tae-kwon-do*, kárate de estilo coreano, me enseñaron a neutralizar los golpes de canto de mano mediante un golpe fuerte de través en el brazo del contrincante. Aquí era importante la desviación del golpe, pero tanto como ello lo era contrarrestar el golpe con una fuerza grande. Así pues, a una fuerza se le enfrentaba otra fuerza y habitualmente ganaba el más robusto. (Rara vez era yo este último.)

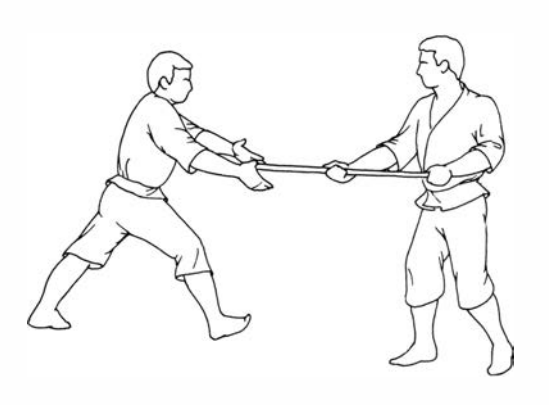
En el aikido el movimiento circular se emplea a la vez para desviar el golpe y para hacer perder el equilibrio al contrario. Supongamos que alguien se nos acerca por detrás, asiéndonos alrededor del cuerpo y aprisionándonos los brazos contra los costados. Entonces, ofreciendo resistencia hacia arriba le mantenemos las manos apretadas firmemente contra nuestro pecho y adelantamos un pie. La coordinación es crítica ya que nuestra intención es desplazar el torso hacia adelante a una velocidad comparable a la del contrario. Si nos retrasamos, perderemos la ventaja de aprovecharnos de su impulso; si nos movemos demasiado rápido, habremos de arrastrarle hacia adelante. Lo que hemos de hacer es deslizar hacia adelante el pie derecho con la celeridad conveniente y después inclinarnos de repente hacia adelante y hacer girar el cuerpo a la derecha.

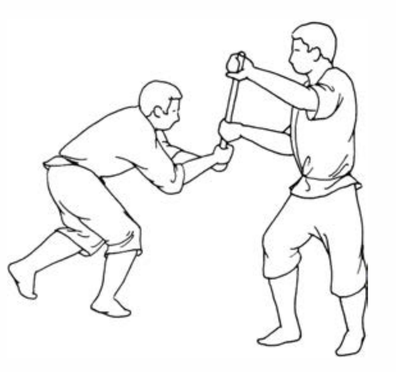
Combinando el impulso inicial del contrario con una rotación propia muy rápida conseguiremos lanzarlo despedido hacia la derecha. En efecto, como nuestra inclinación hacia adelante habrá desplazado su centro de masa por delante de los pies no le será posible evitar nuestro movimiento; como le hemos sujetado las manos y él nos tiene asidos alrededor de los brazos, no le será posible liberarse del movimiento hacia adelante establecido a consecuencia de su impulso. Entonces, la fuerza centrífuga actuante sobre él durante nuestra rápida rotación será excesiva para que pueda aguantarla, considerado lo inestable de su posición, y, por ello, ineludiblemente dará contra el suelo.

En dos ejemplos más de cómo en el aikido se emplea una fuerza pequeña para hacer perder el equilibrio a un atacante aparece la lucha con bastones, tal como se enseña en las clases superiores. Supongamos que un atacante nos lanza un golpe de bastón dirigido al centro de nuestro cuerpo, avanzando su pie derecho durante la acometida y blandiendo el bastón horizontalmente, sosteniéndolo con las palmas de ambas manos hacia abajo. Sería vano tratar de



Cómo luchar co



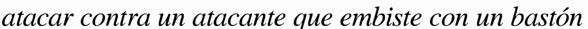


Cómo actuar c





atacar contra un atacante que embiste con un bastón



detener el bastón por el extremo. Lo que haremos es dar un paso atrás con el pie derecho de forma que el bastón pase junto a nuestra izquierda. (La agilidad para ello tan sólo puede proceder de una larga práctica.) Luego, a la vez que el bastón pasa, giramos nuestro cuerpo para quedar frente al mismo y de forma que podamos asirlo con ambas manos. Estas deben situarse de tal manera que la mano izquierda quede delante de la mano más adelantada del atacante y la mano derecha entre las dos manos del atacante.

Al asir el bastón, nuestro propósito no es detener su movimiento, lo que exigiría una fuerza considerable. En lugar de ello lo que se pretende es desviar el bastonazo hacia arriba, haciéndole describir un círculo en torno a nuestra izquierda y después hacia arriba y sobre la cabeza del atacante, pues una vez que éste se ha enzarzado en la acometida hacia adelante poco podrá hacer para evitar la desviación. Efectivamente, para detener su impulso el atacante necesitaría una fuerza considerable y no podrá lanzar un golpe horizontal al centro de nuestro cuerpo mientras tira hacia abajo para impedir nuestra acción de desviación.

Una vez el bastón por encima de la cabeza del atacante, a éste se le derriba fácilmente. Con el pie izquierdo adelantado se encontrará en una posición de gran inestabilidad frente a toda tracción que se le ejerza hacia la izquierda de su retaguardia, ya que en tal dirección su centro de masa tan sólo necesita desplazarse una corta distancia para encontrarse fuera de la superficie de apoyo. Cuando el bastón esté encima de su cabeza, tiraremos hacia abajo y hacia su espalda en dicha dirección, con lo que caerá sobre la colchoneta y probablemente suelte el bastón.

Supongamos ahora que nosotros tenemos el bastón y que un atacante determinado se abalanza hacia adelante para asirlo por su extremo exterior. Le dejaremos que lo agarre, pero guiándole (como una zanahoria delante de un asno) de forma que continúe su embestida. Además alzaremos nuestro extremo del bastón para engañarle y conseguir que se doble hacia adelante. Una vez que le tengamos entregado a este torpe movimiento y esté a punto de pasar por nuestra derecha, levantaremos nuestro extremo del bastón a la altura de su cara y después lo haremos descender hacia su espalda. Si este movimiento se ejecuta con rapidez, le sorprenderemos cuando todavía tiene asido con fuerza el bastón y, por tanto, se doblará hacia atrás a causa de nuestra tracción hacia su espalda y del movimiento continuado de su cuerpo hacia adelante. Así, el momento de giro producido por su propio cuerpo lo hará girar en dirección al suelo en torno a un centro situado a sus pies. Además, la fuerza que ejerce al asir el bastón nos proporciona otro momento de giro para hacerlo girar. De hecho, será él mismo quien provoque su derribo sobre la colchoneta, a causa de su embestida inicial hacia adelante y de un poco de maña por nuestra parte al desestabilizarlo.

El aikido posee centenares de técnicas para utilizar mañosidades como éstas frente a un contrincante determinado. En casi todas ellas una pequeña fuerza de desviación sirve para evitar la acometida del contrario y guiarle de forma que sea él quien se derribe a sí mismo. Siempre que contemplo una defensa de un maestro de aikido, los movimientos me parecen tan fluidos y sin esfuerzos que me veo inclinado a sospechar que el contrario finge cuando cae al suelo. Pero la caída no es simulada; lo parece porque el maestro ha invertido años en desarrollar una sensibilidad intuitiva acerca de las leyes físicas relativas a las fuerzas, los movimientos de rotación y los momentos de giro.





actuar cuando el contrario intenta asir el bastón propio

Libros

Dos siglos de ciencia y técnica, planetología y el incierto futuro del mar Mediterráneo

Arnau, Ramón Margalef y Enrique Macpherson

CIENCIA Y TÉCNICA EN LA SOCIEDAD ESPAÑOLA DE LOS SIGLOS XVI Y XVII, por José María López Piñero. Editorial Labor, Barcelona 1980. Cuando en el mercado aparece una obra seria sobre la historia de la ciencia y la técnica siempre es bien recibida, por cuanto se trata de un tema que ha merecido muy escasa atención por parte de los historiadores españoles y que "no ha iniciado todavía su institucionalización en nuestro país". "Es por ello por lo que se producen confusiones básicas que en otros ambientes resultarían pintorescas, la más grave de las cuales es el escamoteo de la autonomía que la historia de la ciencia tiene como disciplina". Si, además, el autor es el profesor López Piñero, uno de los escasos historiadores que ha logrado crear un grupo de profesionales dedicado a la historia de la ciencia, en Valencia, con una labor realizada hasta el momento verdaderamente notable, entonces la satisfacción por la lectura que nos aguarda es doble: por el tema y por la seguridad de encontrar en ella, en mayor o menor extensión, incentivos diversos para el estudio y la reflexión.

Y así me ha sucedido con la lectura de "Ciencia y técnica en la sociedad española en los siglos XVI y XVII", escrita de acuerdo con la idea comprehensiva que Pierre Vilar acuñara en la expresión "historia total". Es evidente que entre las actividades humanas más variadas –producción, lucha de clases, poder político, comercio, economía en general y un largo etcétera–, la ciencia y la técnica, que, en determinados momentos, han contribuido de una manera decisiva a cambiar la faz de la humanidad, no pueden olvidarse. Y eso es bueno considerarlo de vez en cuando. Es bueno leerlo. Hoy más que nunca, pues, nos encontramos en una encrucijada en la que se cuestiona la ciencia misma.

López Piñero expone los resultados de su estudio en dos partes consecutivas: en la primera, examina la inserción de las actividades científicas y técnicas en la sociedad de los siglos XVI y XVII. Con profundidad, sin perder amenidad, aborda distintas cuestiones: qué se entendía por ciencia y por técnica, cómo se relacionaban (cómo iba desapareciendo la escisión tajante de la Edad Media), cuál era la posición social de quienes producían tales actividades y, por último, la organización de la ciencia y la técnica: instituciones, textos, difusión. Llega por este camino al capítulo IV, el más extenso, en el que se tratan todas las disciplinas cultivadas que previamente él mismo había clasificado: la matemática, la cosmología y astrología –punto aparte merece el copernicanismo–, el arte de navegar, la geografía, la filosofía natural –entendiendo por tal lo que hoy conocemos por física–, la ingeniería, el beneficio de los minerales, sin olvidar las minas americanas, el ensayo de los metales y la destilación, secuela inherente a las últimas materias citadas que constituían lo que hoy llamaríamos ciencia química.

Hemos de resaltar que es en esta parte donde el autor usa a menudo el método dialéctico, tratando sucesivamente la interrelación ciencia-técnica (conocimiento-aplicación); renovación-tradición, ciencia-para-unos-pocos - difusión de la ciencia; método que le permite ofrecer una visión verdaderamente rica de lo que fueron y representaron la ciencia y la técnica de los siglos estudiados.

En la segunda parte estudia las condiciones en que se produjo, en el XVII, el corte entre España y Europa y la posterior voluntad de restablecer el contacto. En otras palabras: el aislamiento y el deseo de reintegración. Distingue para ello, en esa centuria, tres etapas: una primera –que se extiende a lo largo del primer tercio del siglo– puede considerarse prolongación del siglo anterior, con cierto prestigio científico aún, especialmente en el campo de la navegación, del beneficio de los minerales, de la ingeniería y la medicina, aunque de espaldas a las novedades que empezaban a invadir el continente; la segunda, de 1630 a 1670, aproximadamente, en la que estas novedades van introduciéndose paulatinamente en la península, pero de forma harto aislada y fragmentaria, siendo aceptadas –cuando lo son– por los científicos españoles como meras rectificaciones del saber anterior o tradicional; en postrer lugar –tercera etapa que dura hasta finales del siglo– la toma de conciencia del atraso científico-técnico. En esta última, rompiendo hasta cierto punto con el saber tradicional, se intenta hacer un esfuerzo por asimilar la ciencia moderna como soporte del período ilustrado que se avecina. Es el momento en que unos pocos comprenden que España ha permanecido fuera del gran movimiento que alumbró la ciencia moderna.

Desearíamos llamar la atención del lector sobre dos aportaciones de sumo interés: una relación de los estudios históricos sobre la actividad científica en la España de los siglos XVI y XVII y un exhaustivo apéndice bibliográfico. La documentación reseñada es una de las partes más sobresalientes de la obra y, a la vez, herramienta indispensable para cualquier estudioso de la cuestión. Resultado de una investigación seria y prolongada, la obra sugiere nuevos temas y tratamientos, mayores profundizaciones y determinaciones concretas.

Hay numerosas tablas, que encierran valiosos datos cuantitativos; las ilustraciones que remiten a tratados de la época hacen agradable su lectura. (A.)

ELS PLANETES COMPARATS, LOS PLANETAS COMPARADOS, COMPARATIVE PLANETARY ENVIRONMENTS, edición al cuidado de J. Oró, J. C. Vergés y A. Plana Castellví. Centre Mediterrani d'Investigació i Desenvolupament, Barcelona, 1980, 235 págs. Este volumen recoge los resultados de un simposio celebrado en Barcelona, los días 9 y 10 de junio de 1979, con las transcripciones de 9 conferencias (5 en catalán, 1 en castellano y 3 en inglés), más los comentarios a que dieron lugar y un coloquio sobre la contaminación, en especial de la Costa Brava. Dentro de esta especie de *pot-pourri* (permítasenos un galicismo para completar su aire internacional) destaca como idea fundamental la importancia que puede tener una planetología comparada para entender mejor el funcionamiento de nuestro propio globo, haciendo ver que una gran parte de sus características dependen de la existencia de la vida –y no a la recíproca, como se tiende a creer–. En este sentido, nuestra Tierra, con su biosfera, adquiere el carácter de unidad evolutiva, con capacidad para regularse a sí misma. Aunque no se diga claramente, este libro está en la

línea de pensamiento divulgada principalmente por Margulis y por Lovelock, que insisten en la unidad funcional de nuestro planeta, y que personalizan con el nombre de "Gaia" o "Gea". No hay duda que esta forma de pensar constituye un marco adecuado donde tratar de situar los problemas de contaminación, recursos y población, que son nuestra preocupación de cada día.

Dentro de estas líneas de guía, los autores y temas de las conferencias principales son los siguientes: Tobias Owen trata de las atmósferas de los planetas (13 págs.), H. P. Klein de la búsqueda de indicios de vida en Marte (12 págs.), Joan Oró del origen de la vida sobre la Tierra (15 págs.), Robert D. MacElroy de la composición de ecosistemas regulables en el laboratorio (6 págs.), Albert Serratosa de los problemas ambientales en las zonas urbanas (13 págs.), Carles Bas de los recursos pesqueros de nuestra plataforma (17 págs.), Tomás Vidal de la distribución de la población humana en Cataluña (14 págs.), Josep C. Vergés de posibles aspectos positivos del factor humano en la regulación de la contaminación fluvial (9 págs.) y Andrés Maldonado traza una perspectiva del origen y evolución del mar Mediterráneo (16 págs.). Todos los autores han mantenido sus textos sin demasiados tecnicismos y al alcance de un público muy amplio, lo cual es una nota muy positiva de este libro, que pretende sacar a los problemas ambientales del nivel ínfimo al que suelen ser debatidos. La ilustración no es extraordinaria; la única lámina en color hubiera podido ser mejor escogida y siento tener que decir que la distribución de la clorofila en las costas saharianas se ilustra con una figura original del autor de este comentario, que no es la apropiada (pág. 120). (R.M.).

EL MEDITERRÁNEO: UN MICROCOSMOS AMENAZADO, por Varios; Editorial Blume; Barcelona 1980. La contaminación y la degradación del medio ambiente por parte del hombre son temas de gran actualidad y que están despertando, tanto en la opinión pública como en los medios científicos y técnicos, una gran expectación. De ahí que cualquier intento de aportar información objetiva sobre el tema sea, en principio, digna de alabanza.

El libro, publicado originalmente por la revista AMBIO de la Real Academia Sueca de las Ciencias, está dividido en 19 capítulos escritos por personas de cierto prestigio internacional: A. M. M'Bow, Director General de la UNESCO; P. Le Lourd, Director Regional del Centro de Estudios sobre la Contaminación por Hidrocarburos; D. Charbonnier, Secretario General del Comité para el Estudio de las Pesquerías del Mediterráneo; R. Helmer, S. Keckes, M. Tangi, P. S. Thacher, del PNUMA. (Programa de las Naciones Unidas para el Medio Ambiente), etcétera.

Por su temática, los 19 capítulos pueden agruparse en varios apartados. Esta agrupación hubiera sido interesante que figurara en la obra a fin de facilitar al lector una mayor continuidad en la lectura. Tal como están estructurados en el texto pueden inducir a un cierto desorden en las ideas. Además de la introducción, habría que distinguir primero una serie de capítulos dedicados a la contaminación propiamente dicha, otros sobre los diferentes acuerdos y planes internacionales que se están llevando a cabo en el Mediterráneo, un tercer grupo sobre ejemplos prácticos de lucha contra la contaminación y degradación del ecosistema y, finalmente, una serie de capítulos en torno a aspectos particulares del mar y características de los países ribereños.

La introducción, bastante general, da una visión amplia del pasado y presente del Mediterráneo y sus límites costeros. El autor, al final del artículo, quizás peque de exagerado al hablar de una hipotética edad de oro de los sistemas rurales y autosuficientes de nuestros antepasados. Indudablemente, la naturaleza estaba menos degradada en siglos pasados y los sistemas actuales inducen a un cierto stress del ecosistema, pero aquellas situaciones también reportaban una serie de inconvenientes que olvidamos fácilmente.

A propósito de los diferentes agentes contaminantes, se describen primero, los que proceden de fuentes terrestres: residuos domésticos, industriales, agrícolas, radiactivos, etcétera; se detalla su origen y su concentración en las diferentes áreas mediterráneas. Al abordar la contaminación por petróleo, se destaca la influencia de las operaciones de lastre y deslastre de petroleros como principal agente contaminante. Se pasa revista a la contaminación bacteriana (*Salmonella, Shigella,* cólera y otros) y sus efectos en la situación sanitaria. Se concluye, a modo de resumen, con el estado actual (con datos hasta 1977) de la contaminación en el mar Mediterráneo. Por los objetivos que parece perseguir el libro, estos capítulos cumplen a la perfección su cometido sin profundizar en pormenores de especialista ni caer en la generalización irrelevante.

El apartado sobre planes de acción y programas piloto consta de otros cuatro capítulos: Plan de acción para el Mediterráneo, Programa coordinado de vigilancia e investigación, Convenio de Barcelona y Plan Azul. La información que en ellos se da es de cierta utilidad ya que pone al corriente de los diferentes proyectos que existen para salvaguardar nuestro mar. No obstante, se peca en todos ellos de un excesivo aire de documento de trabajo o de resumen de las conclusiones de una asamblea, lo que en algunas ocasiones hace que su lectura resulte algo espesa.

Al detallar los ejemplos prácticos en la lucha contra la degradación del medio ambiente, analiza el fenómeno del turismo y sus efectos, resaltando la necesaria planificación del mismo. Se detiene en la ciudad de Haifa, ejemplo típico de población industrializada y con escasa programación urbanística. Son de gran interés las medidas que ha tomado el gobierno israelí y el ayuntamiento de la ciudad (calidad ambiental, control de residuos, vertidos, contaminación en playas, etcétera). Y hace escala también en Atenas y las medidas de control de la contaminación allí tomadas, en particular para el caso de la Acrópolis. Se cierra el apartado con dos interesantes textos sobre la lucha contra la desertización en toda la región mediterránea y en Túnez en particular, desvelándose los factores que influyen en estos procesos (el climático, el humano y el económico).

Queda un cuarto y último apartado que engloba aspectos peculiares del ecosistema mediterráneo. Entre ellos, la degradación de la maquia, quizás uno de los más interesantes de toda la obra. El artículo está tratado con bastante profundidad y rigor científico. El análisis, pormenorizado, saca a luz los distintos factores que influyen en este tipo de vegetación (ambientales, dinámicos, ecológicos, económicos, etcétera). Se trata del artículo mejor estructurado del libro.

Al considerar las reservas de agua dulce en la cuenca mediterránea señala la distribución y abundancia de las mismas y las medidas encaminadas a un mejor aprovechamiento. Se describen someramente los recursos pesqueros y las perspectivas existentes. Aunque aporta abundante información, y muy bien estructurada, le falta profundidad y quizás es excesivamente optimista en lo concerniente al futuro de la acuicultura. Hubiera sido deseable, por otra parte, una mayor información sobre los beneficios que se obtendrían con una explotación más racional. (E.M.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

LA TRADUCCION DE LOS LENGUAJES DE COMPUTADOR

COMPILER CONSTRUCTION. AN ADVANCED COURSE. Dirigido por G. Goos y J. Hartmanis. Springer-Verlag, 1974.

APL-AN INTERACTIVE APPROACH. Leonard Gilman y Allen J. Rose. John Wiley and Sons, Inc., 1974.

IL. AN INTERMEDIATE SYSTEMS PROGRAMMING LANGUAGE. María Luisa Tavera y Manuel Alfonseca. Publicación del Centro de Investigación UAM-IBM n.° SCR 01-78; mayo, 1978.

A MACHINE INDEPENDENT APL INTERPRETER. Manuel Alfonseca y María Luisa Tavera en *IBM Journal of Research and Development,* vol. 22, n.° 4, págs. 413-421; julio, 1978.

LUPUS ERITEMATOSO SISTEMICO

IMMUNOPATHOGENESIS OF SYSTEMIC LUPUS ERITHEMATOSIS. David Koffler en *Annual Review of Medicine,* vol. 25, págs. 149-164; 1974.

THE IMMUNOLOGIC APPROACH TO SYSTEMIC LUPUS ERYTHEMATOSUS. H. G. Kunkel en *Arthritis and Rheumatism,* vol. 20, págs. S139-S147; 1977.

RHEUMATOID DISEASES. Ken Fye, Haralampos Moutsopoulos y Norman Talal en *Basic and Clinical Immunology,* dirigido por H. Hugh Fudenberg, D. P. Stites, J. L. Caldwell y J. V. Wells. Lange Medical Publications, 1978.

RECENT PROGRESS IN THE STUDY OF AUTOANTIBODIES TO NUCLEAR ANTIGENS. R. M. Nakamura y E. M. Tan en *Human Pathology,* vol. 9, n.° 1, págs. 85-91; enero, 1978.

THE IMMUNOLOGY OF RHEUMATOID DISEASES. David Koffler en *CIBA Clinical Symposia,* vol. 31, n.° 4; 1979.

ASTRONOMIA DE LINEAS DE RAYOS GAMMA

GAMMA-RAY ASTRONOMY. E. L. Chupp. D. Reidel Publishing Co., 1976.

EVIDENCE FOR STRONG CYCLOTRON EMISSION IN THE HARD X-RAY SPECTRUM OF HERCULES X-1. J. Trümper, W. Pietsch, C. Reppin, W. Voges, R. Staubert y E. Kendziorra en *The Astrophysical Journal Letters,* vol. 219, n.° 3, parte 2, págs. L105-L110; 1 de febrero de 1978.

GAMMA-RAY LINES: A NEW WINDOW TO THE UNIVERSE. Richard E. Lingenfelter y Reuven Ramaty en *Physics Today,* vol. 31, n.° 3, págs. 40-47; marzo, 1978

MONUMENTOS MEGALITICOS

STONEHENGE, R. J. C. Atkinson. Macmillan and Company, 1956.

THE MEGALITH BUILDERS OF WESTERN EUROPE. Glyn L. Daniel. Praeger Publishers, 1958.

BRITTANY. P. R. Giot. Praeger Publishers, 1960.

ANCIENT EUROPE FROM THE BEGINNINGS OF AGRICULTURE TO CLASSICAL ANTIQUITY. Stuart Piggott. Edinburgh University Press, 1965.

BEFORE CIVILIZATION: THE RADIOCARBON REVOLUTION AND PREHISTORIC EUROPE. Colin Renfrew. Alfred A. Knopf, Inc., 1973.

CONVECCION

HYDRODYNAMIC AND HYDROMAGNETIC STABILITY. S. Chandrasekhar. Oxford University Press, 1961.

BUOYANCY EFFECTS IN FLUIDS. J. S. Turner. Cambridge University Press, 1973.

CONVECTIVE INSTABILITY: A PHYSICIST'S APPROACH. Christiane Normand, Yves Pomeau y Manuel G. Velarde en *Reviews of Moderns Physics,* vol. 49, n.° 3, págs. 581-624; julio, 1977.

TALLER Y LABORATORIO. Jearl Walker en *Investigación y Ciencia,* n.° 15, págs. 116-121; diciembre, 1977.

FISIOLOGIA DEL KOALA

THE KOALA: PROCEEDINGS OF THE TARONGA SYMPOSIUM ON KOALA BIOLOGY, MANAGEMENT, AND MEDICINE. Dirigido por T. J. Bergin. Zoological Parks Board of N.S.W., 1978.

METABOLISM AND HEAT BALANCE IN AN ARBOREAL MARSUPIAL, THE KOALA (*PHASCOLARCTOS CINEREUS*). Robert Degabriele y T. J. Dawson en *Journal of Comparative Physiology.* vol. 134, págs. 293-301; 1979

ADN SUPERENROLLADO

STRUCTURE AND REACTIONS OF CLOSED DUPLEX DNA. William R. Bauer en *Annual Review of Biophysics and Bioengineering,* vol. 7, págs. 287-313; 1978.

DNA TOPOISOMERASES: ENZYMES THAT CATALYZE THE CONCERTED BREAKING AND REJOINING OF DNA BACKBONE BONDS. James C. Wang en *Molecular Genetics,* dirigido por J. Herbert Taylor. Academic Press, 1979.

IS DNA REALLY A DOUBLE HELIX? F. H. C. Crick J. C. Wang y W. R. Bauer en *Journal of Molecular Biology,* vol. 129, n.° 3, págs. 449-457; 15 de abril de 1979.

DNA GYRASE AND THE SUPERCOILING OF DNA. Nicolas R. Cozzarelli en *Science.* vol. 207, n.° 4434, págs. 953-960; 29 de febrero de 1980.

LA HELICE

AIRPLANE PROPELLERS. Herman Glauert en *Aerodynamic Theory,* dirigido por William Frederick Durand. Springer-Verlag, 1934-36.

AERODYNAMICS: SELECTED TOPICS IN LIGHT OF THEIR HISTORICAL DEVELOPMENT. Theodore Von Kármán. Cornell University Press, 1954.

MARINE HYDRODYNAMICS. J. N. Newman. The MIT Press, 1977.

JUEGOS MATEMATICOS

AN EPISODE OF FLATLAND. C. H. Hinton. Swan Sonnenschein & Co., 1907.

PLANILANDIA. E. A. Abbott. Ediciones Guadarrama. Madrid, 1976.

SPHERELAND: A FANTASY ABOUT CURVED SPACES AND AN EXPANDING UNIVERSE. Thomas Y. Crowell, 1965.

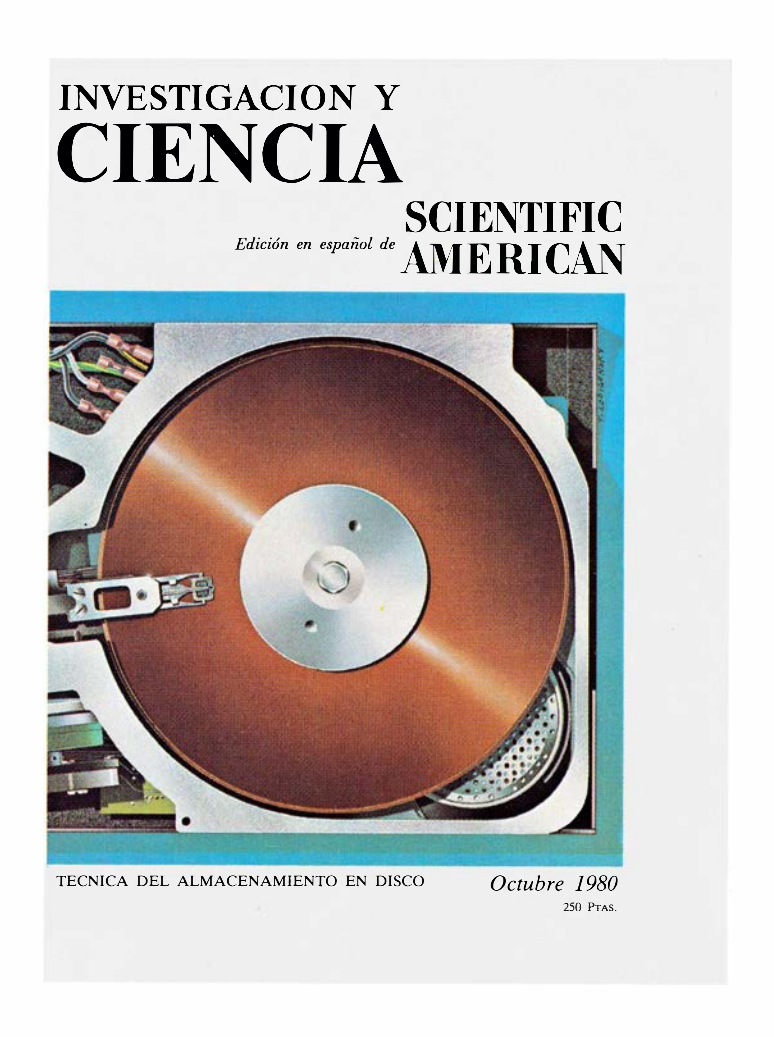
TALLER Y LABORATORIO

JUDO ON THE GROUND. E. J. Harrison, W. Foulsham & Co., 1954.

AIKIDO AND THE DYNAMIC SPHERE: AN ILLUSTRATED INTRODUCTION. A. Westbrook y O. Ratti. Charles E. Tuttle Company, 1970.

ADVANCED TECHNIQUES. Morihiro Saito en *Traditional Aikido: Sword, Stick and Body Arts.* Minato Research & Publishing Co., 1973-74.

Seguiremos explorando los campos del conocimiento



TECNICA DEL ALMACENAMIENTO EN DISCO, por Robert M. White
En la superficie de un disco giratorio los datos para un ordenador se almacenan como regiones magnéticas diminutas. En los métodos que se están desarrollando, un láser "escribirá" y "leerá" los datos.

LA MANZANA DE NEWTON Y EL DIALOGO DE GALILEO, por Stillman Drake
Newton vio caer la manzana, pero ¿qué fue lo que le inspiró la relación entre la caída de la manzana y la órbita lunar y llegar luego a la ley de la gravedad universal? Un diagrama del "Diálogo".

MICROTUBULOS, por Pierre Dustin
Son orgánulos estructurales que existen en todas las células nucleadas. Ensamblados como un andamiaje a partir de subunidades proteicas, participan en la división, el movimiento y el mantenimiento de la forma celular.

LA SUPERFICIE DE VENUS, por Gordon H. Pettengill, Donald B. Campbell y Harold Masursky
Oculta por nubes, se ha cartografiado ahora por radar desde la Tierra y desde un vehículo espacial en órbita alrededor de Venus. Las imágenes sugieren una constitución geológica intermedia entre la de la Tierra y la de Marte.

EL ELECTRON AISLADO, por Philip Ekstrom y David Wineland
Mediante un único electrón atrapado en un átomo artificial de tamaño macroscópico, se puede medir con precisión no superada una propiedad del electrón llamada el factor g.

LOS ANTECEDENTES DE LA CIVILIZACION EN EL VALLE DEL INDO, por Jean-François Jarrige y Richard H. Meadow
Una de las tres civilizaciones más antiguas fue la del Indo. Las excavaciones realizadas en Mehrgarh, Pakistán, han descubierto poblados de labriegos que florecieron en aquella región 3000 años antes.

LA PLACENTA, por Peter Beaconsfield, George Birdwood y Rebecca Beaconsfield
Este extraordinario órgano actúa de intermediario entre dos organismos diferentes: el feto y la madre. Después del parto, puede utilizarse como "animal de experimentación" para una gran variedad de estudios.

LAS AREAS OCEANICAS MAS PRODUCTIVAS, por Ramón Margalef y Marta Estrada
Las áreas muy ricas en pesca, que habían hecho concebir esperanzas excesivas en los océanos como "reservas inagotables de proteínas para la humanidad", son muy limitadas; en realidad, la mayor parte de la extensión de los mares es comparable a un desierto.

